THE ALBERTA JOURNAL OF EDUCATIONAL RESEARCH

VOLUME XXXIV NUMBER 4 DECEMBER 1988

PUBLISHED BY
THE UNIVERSITY OF ALBERTA • EDMONTON

THE ALBERTA JOURNAL OF EDUCATIONAL RESEARCH

Published in March, June, September, and December by
the Faculty of Education, University of Alberta

A quarterly journal devoted to the dissemination, criticism, interpretation, and encouragement of all forms of systematic enquiry into education and fields related to or associated with education.

EDITOR: Walter H. Worth

CONSULTING EDITORS

Reno A. Bosetti,
Alberta Department of Education

Robert C. Bryce,
University of Regina

Janet A. Donald,
McGill University

Michael G. Fullan,
University of Toronto

Jane Gaskell,
University of British Columbia

Myrna L. Greene,
University of Lethbridge

Lilian G. Katz,
University of Illinois

Joanne C. Keselman,
University of Manitoba

John McLeod,
University of Saskatchewan

Alan R. Pence,
University of Victoria

Lauran H. Sandals,
University of Calgary

Hugh A. Stevenson,
University of Western Ontario

Keith C. Sullivan,
Dalhousie University

Jaap Tuinman,
Simon Fraser University

John D. Wilson,
Moray House College of Education

FACULTY ADVISORY COMMITTEE

Thomas O. Maguire (Chair)
Alton T. Olson
Erwin Miklos
Fern Snart

AJER gratefully acknowledges support from the Social Sciences and Humanities Research Council of Canada and from the Alberta Advisory Committee for Educational Studies.

The subscription rate is $32.00 per year for individuals, $40.00 per year for institutions. Add $8.00 for delivery outside Canada. Single copies are $8.00 each. Please make cheques payable to *The Alberta Journal of Educational Research*. All back issues are available; rates supplied on request. Claims for undelivered copies must be received within three months of publication.

Address all communications and manuscript submissions to *The Alberta Journal of Educational Research*, Faculty of Education, Publication Services, 4-116 Education North, University of Alberta, Canada, T6G 2G5.

SECOND CLASS MAIL REGISTRATION NUMBER 1436

Faculty of Education
University of Alberta

ISSN 0002-4805

The Alberta Journal of Educational Research

Volume XXXIV, Number 4 *December 1988*

CONTENTS

THE EDITOR'S PAGE: Rules for Reviewers 331

Social Psychological Dispositions and Academic Achievement of Inuit and Non-Inuit Students 332
R.A. Clifton and L.W. Roberts

Reading Processes Subgroups in a Clinical Population 344
G.V. Malicky and C.A. Norman

Exploratory Analysis of Disagreement Among Holistic Essay Scores 355
P. Nagy, P.J.A. Evans, and F.G. Robinson

Leaders' Priorities and the Congruity Imperative 375
J. Fris and J.H. Balderson

Perceived Attributes of Mainstreaming, Principal Change Strategy, and Teacher Attitudes Toward Mainstreaming 390
C. Duquette and R.R. O'Reilly

PERSPECTIVES

Policy Research and Politics 403
J. Gaskell

BOOK REVIEW

Introduction to the Economics of Canadian Education by D.M. Richards and E.W. Ratsoy 418
Reviewed by D.J. Brown

The Alberta Journal of Educational Research *Vol. XXXIV, No. 4, December 1988, 331-331*

THE EDITOR'S PAGE

Rules for Reviewers

This issue marks the end of my four-year term as Editor. That I found the preparation of the 16th issue as interesting and as enjoyable as the first one is attributable to the splendid cooperation and assistance received from so many people—authors, the staff of Publication Services, consulting editors, members of the Faculty Advisory Committee, and those unsung heroes, our reviewers.

Each year about a hundred different persons act as reviewers of manuscripts submitted to *AJER*. They undertake this professional responsibility willingly and perform it admirably. They serve anonymously without compensation. Their satisfaction comes from helping others in advancing knowledge of the educational enterprise. Their reward for doing a good job is to be asked to do it again. Reflecting on my experience with the peer review process and the dedicated people who make it work leads me to propose some rules for prospective reviewers.

Be thorough. Read the entire manuscript to sense its logic and flow. Then look for answers to questions like these: Is the rationale of the study apparent and the purpose clearly stated? Are the data sources, data collection, and data analysis procedures described and appropriate? Are the findings and conclusions unambiguous, valid, and meaningful? Is the relationship of the investigation to the work of others clarified? Report your answers to such questions.

Consider significance. Always comment on the theoretical and/or practical importance of the research question or topic. What kind of impact is it likely to have on the activities, behavior, or work of others?

Highlight strengths. Few manuscripts have no redeeming features. Find the good points and mention them. Most authors will undertake to revise their work when they believe they are building on strength.

Offer advice. Help the author ready the manuscript for publication by providing suggestions for additions, changes, and deletions. In addition to editorial revisions, you may be able to identify some appropriate statistical procedure. Similarly, you may be able to put forward an alternative or expanded interpretation of the findings. Authors are usually very appreciative of and quite responsive to the advice received.

But the most important rule of all for reviewers to remember is the *golden rule*—do unto others as you would have them do unto you.

This latter rule also applies to editors. Because my successor Tom Maguire practiced this in his previous role as Associate Dean (Graduate Studies and Research), I am confident *AJER* will prosper under his editorship.

W.H.W.

The Alberta Journal of Educational Research *Vol. XXXIV, No. 4, December 1988, 332-343*

RODNEY A. CLIFTON

and

LANCE W. ROBERTS
University of Manitoba

Social Psychological Dispositions and Academic Achievement of Inuit and Non-Inuit Students

This study examines the differences between Inuit and non-Inuit students in activism, self-concept of ability, and academic achievement. Activism and self-concept of ability are designated as intervening variables between ethnicity and academic achievement. The results show that Inuit students have lower activism and self-concept of ability scores than non-Inuit students. Moreover, these two social psychological dispositions account for a large part of the effect ethnicity has on academic achievement. We suggest that these results have implications for the teachers of Inuit students. Specifically, this evidence lends support to Kleinfeld's (1975) suggestion that effective teachers of Inuit students create emotionally warm and personalized classroom environments, while demanding high academic achievement. In doing so, teachers may positively affect students' activism and self-concept of ability which, in turn, may enhance academic achievement.

A persistent theme concerning the education of Inuit students is their relative lack of success. Compared to non-Native students, Inuit students generally have lower achievement on standardized tests, are usually behind in their age-grade placement, and are more likely to drop out of school (Clifton, 1972, 1977; Finley, 1983; Honigmann & Honigmann, 1970, pp. 176-205; Kleinfeld, 1975, 1979; Vernon, 1966, 1984).

The limited research literature that is available suggests that social psychological factors account, to some degree, for the underachievement of Inuit students. This literature suggests that Inuit students are probably treated differently from non-Inuit students by their teachers, and this leads to their underachievement. Kleinfeld (1975), for example, reports that this process is similar to that illustrated in the classic study of formal education among the Sioux by Wax, Wax, and Dumont (1964). In this case, the teachers routinely disparaged the Sioux culture because they assumed that their mission was to reform the students by teaching the manners of white society as moral absolutes. The Sioux students responded, both psychologically and sociologically, to this type of treatment. The deprecation of their culture affected the students' psyches and was evident in low self-concepts

---

Dr. Clifton is a member of the Department of Educational Administration and Foundations, and his interests are in the areas of ethnicity and teacher education.

Dr. Roberts is with the Department of Sociology with interests in ethnicity and social psychology.

of ability. The students' social response took the form of group solidarity based on passive resistance. They formed a tightly integrated group that refused to actively participate in classroom work. The result is what many observers have reported—shy, withdrawn, self-effacing, nonparticipating Native students who individually display a *mask of silence* and who collectively constitute the *silent classroom* (Bienvenue, 1978; Clifton, 1977; Franklyn, 1974; Hawthorn, 1967; Kleinfeld, 1979; Rampaul, Singh, & Didyk, 1984; Wax, Wax, & Dumont, 1964). In short, the literature suggests that differences between Inuit and white students are displayed in terms of the students' self-concept of ability and their activism. This implies that Inuit students generally have lower assessments of their potential and have less active dispositions toward education, which together produce lower academic achievement.

At the same time, it may be suggested that the differences between Inuit and white students may reflect differences in socioeconomic status rather than in culture. This seems plausible because Native people generally have low socioeconomic status (Clifton, 1977; Fuchs & Havighurst, 1973; Hawthorn, 1967; Taylor & Skanes, 1976). Thus it is relevant to take socioeconomic status into account when examining the effects of ethnicity on self-concept of ability, activism, and academic achievement. Similarly, some studies suggest that the sex of the students, as well as the length of time the students have been in school, must be controlled (Clifton, 1975; King, 1967; Wolcott, 1967). It is suggested that the adverse effects of ethnicity on self-concept of ability, activism, and academic achievement may differ by sex and may increase as Native students spend longer periods of time in unsupportive educational environments (Barnes & Vulcano, 1982; Clifton, 1975, 1977; Hawthorn, 1967; King, 1967; Rampaul, Singh, & Didyk, 1984; Toohey, 1985a, 1985b; Wolcott, 1967).

From this research literature we have developed a theoretical model which guides our study (Figure 1). In this model, ethnicity, sex, years of school, and socioeconomic status are designated as independent variables; activism and self-concept of ability are designated as intervening variables; and academic achievement is designated as the dependent variable. This exploratory study expands on previous research by taking into account several relevant variables such as socioeconomic status, sex, and years of school that have been considered in isolation in previous studies. As a result, we can distinguish, for example, between the effects of ethnicity and the effects of socioeconomic status on activism and self-concept of ability and academic achievement of Inuit students. Moreover, we can distinguish between the direct and the indirect effects that ethnicity has on academic achievement.

Methods

Variables

Data on all seven variables included in this model were collected from a survey questionnaire and an achievement test completed by students in two northern schools. Three of the independent variables—ethnicity, sex, and years of school—are coded in a straightforward manner. Ethnicity is coded

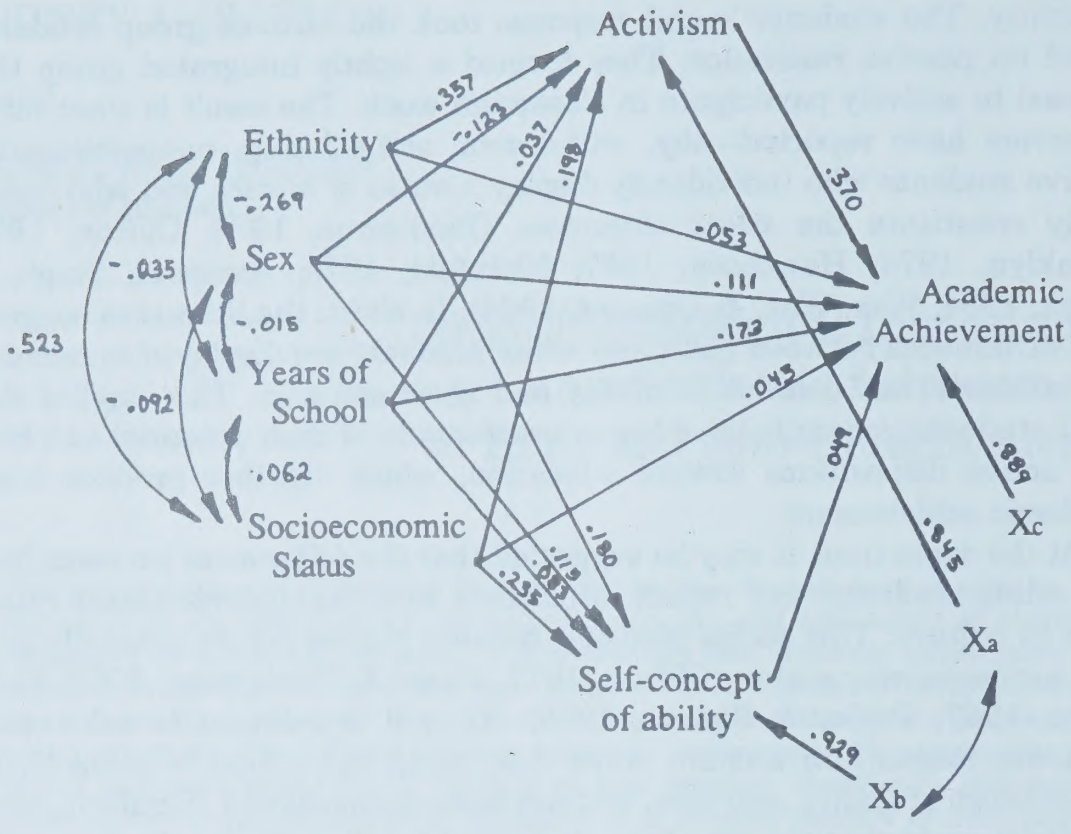

Figure 1: The path model representing the strength of the relationships between the independent, intervening, and dependent variables

as 1 for Inuit and 2 for non-Inuit, sex is coded as 1 for males and 2 for females, and years of school is coded as the actual number of years a student had been enrolled. The other four variables—academic achievement, activism, self-concept of ability, and socioeconomic status—have the following more complex operationalizations.

Academic achievement. Constructing a measure of academic achievement is problematic, particularly because it is difficult to separate the effects of English competence from academic achievement. Several studies report that Inuit and other Native students score lower on English verbal ability tests than non-Native students (Clifton, 1977; Kleinfeld, 1979; MacArthur, 1973, 1978; Marjoribanks, 1972; Mueller, Mulcahy, Wilgosh, Watters, & Mancini, 1986; Toohey, 1985a, 1985b; Vernon, 1966, 1984). In particular, these studies suggest that Inuit students have difficulty understanding and participating in classroom activities because they lack English competence, which contributes to their high failure and drop-out rates. Nevertheless, English has, and continues to be, the principal language of instruction and evaluation in most northern schools (Toohey, 1985b). For the purposes of this study, English competence and academic achievement are combined into a measure of academic achievement.

The academic achievement scale is derived from a test of 40 word pairs which were arranged so that students had to indicate whether they were synonyms or antonyms. The test was originally developed by the Interna-

tional Association for the Evaluation of Educational Achievement (Comber & Keeves, 1973, pp. 393-394; Thorndike, 1973, p. 17) and has been administered to samples of 10-year-old students in 15 countries. The Kuder-Richardson reliability coefficient of the test is .84, which is similar to the reliability coefficients reported for students in other countries (Thorndike, 1973, p. 58).

Activism. This scale was derived from Kahl's (1965, 1968) and Inkeles' (Inkeles & Smith, 1977) instruments to measure social psychological dispositions related to modernity and refers to planning for a predictable future (Roberts, Clifton, & Wiseman, in press). This scale was composed of four items: Planning only makes a person unhappy, since your plans hardly ever work out; Making plans only brings unhappiness because the plans are hard to fulfill; With things as they are today, an intelligent person ought to think only about the present, without worrying about what will happen tomorrow; and It doesn't make much difference whom people elect to government, for nothing will change. These items were coded on five-point scales from *completely agree* to *completely disagree*. Item-total correlations range from .26 to .48, and the alpha reliability coefficient for the scale is .61.

Self-concept of ability. This scale was derived from Brookover's Self-Concept of Ability Scale (Brookover, Beady, Flood, Schweitzer, & Wisenbaker, 1979, p. 207). The scale, as used in this study, is composed of four items: I feel that I am as able as my best friends to get good marks; I feel that I am able to go on to university; I feel that I am as able as my classmates to get good marks in school; and I feel that I am able to complete high school. These items were coded on five-point scales from *completely disagree* to *completely agree*. Item-total correlations ranged from .27 to .44, and the alpha reliability coefficient for the scale is .57.

Socioeconomic status. In order to measure socioeconomic status, the students were asked to respond to a question about their fathers' occupations. The responses were coded so that scores below 40.68 on the Blishen and McRoberts (1976) scale were coded as 1, and scores at or above 40.68 were coded as 2. The sample was dichotomized into blue-collar and white-collar categories because the distribution was skewed. Approximately 67% of the sample had fathers in blue-collar occupations.

Sample

Students participating in this study were enrolled in the public schools of two communities in the eastern part of the Northwest Territories. The total population of students attending these two schools, 46 Inuit and 11 non-Inuit students, were included in this study. The Inuit students ranged in age from 10.5 to 16.5 years and had attended school for between three and 11 years. More than 50% of them had spent either six or seven years in school. The Inuit students were quite evenly divided between the sexes, with 24 females and 22 males in the sample. The non-Inuit students ranged in age from just under 11 years to just under 16 years. The time that they had spent in school varied between five and 10 years, and over 50% of them had

been at school for between six and seven years. The non-Inuit students included two females and nine males. In sum, the Inuit and non-Inuit students were similar, except for their sex distributions.

The fact that the data for this study are from a small, nonrandom sample of students clearly limits the generalizability of the results. However, small samples have a legitimate place in social research, especially for exploratory studies in areas in which there has been a lack of research (Mueller, Schuessler, & Costner, 1977, p. 370). Moreover, discussions with other researchers and teachers suggest that these two schools are typical of integrated schools throughout northern Canada. On balance, a cautious use of our sample to form some preliminary conclusions seem warranted.

Results

The correlation matrix, means, and standard deviations for the seven variables are presented in Table 1. These correlations were computed on the basis of pair-wise deletion of missing data and were used to estimate the parameters of the model using ordinary least squares regression procedures. In this table, it is evident that ethnicity is moderately correlated with all other variables in the model except for years in school. It is also evident that academic achievement is moderately correlated with socioeconomic status, activism, self-concept of ability, and, not surprisingly, with the number of years a student has been enrolled in school.

In Table 2 the standardized and unstandardized regression coefficients are reported for the effects of the independent variables on the two intervening variables, activism and self-concept of ability, and the dependent variable, academic achievement. In the standardized form each variable has been transformed so that it has a mean of zero and a standard deviation of one, while in the unstandardized form each variable has a mean and a standard deviation which reflect the metric properties of the variable as

Table 1

Correlation Matrix, Means, and Standard Deviations for the Variables in the Theoretical Model

	1	2	3	4	5	6	7
1. Ethnicity							
2. Sex	−.269						
3. Years of school	−.035	−.015					
4. Socioeconomic status	.523	−.092	.062				
5. Activism	.493	−.238	.039	.399			
6. Self-concept of ability	.269	.044	.090	.323	.029		
7. Academic achievement	.236	.004	.191	.237	.396	.109	
Means	1.19	1.46	7.11	1.33	12.00	15.72	13.09
SDs	0.40	0.50	1.56	0.47	3.46	2.68	6.08

Table 2
Standardized and Unstandardized Regression Coefficients and R^2s for the Intervening and Dependent Variables[a]

	Intervening Variables		Dependent Variable	
Independent Variables	Activism	Self-concept of Ability	Academic Achievement	Academic Achievement
Ethnicity	.357 (3.108)	.180 (1.213)	.194 (2.968)	.053 (.815)
Sex	−.123 (−.849)	.115 (.614)	.071 (.862)	.111 (1.346)
Years of school	.037 (.082)	.083 (.143)	.191 (.745)	.173 (.674)
Socioeconomic status	.198 (1.451)	.235 (1.331)	.130 (1.666)	.045 (.576)
Activism				.370 (.650)
Self-concept of ability				.049 (.111)
R^2	.286	.137	.113	.209

[a]Unstandardized coefficients are in parentheses.

reported in Table 1. For the dependent variable, regression coefficients are reported for both reduced-form (column 3) and fully recursive (column 4) analyses. As such, both the total and indirect causal effects of the independent variables on academic achievement can be calculated (Alwin & Hauser, 1975). Specifically, the total causal effects are the coefficients reported in column 3, the direct effects are the coefficients reported in column 4, and thus the indirect effects, as mediated by activism and self-concept of ability, are computed by subtracting the relevant coefficients in column 4 from the relevant coefficients in column 3.

In Figure 1 the standardized regression coefficients are reported as path coefficients representing the relationships between the independent, intervening, and dependent variables. The coefficients on the curved lines between the independent variables are the correlation coefficients as reported in Table 1, and the coefficients on the lines originating with Xs are the effects of the residuals. This model is congruent with the established conventions of representing path models (Duncan, 1975; Heise, 1975).

Ethnicity has relatively large effects on the two social psychological variables, activism (.357) and self-concept of ability (.180). In both cases, non-Inuit students have higher scores than Inuit students. Nevertheless, the effect of ethnicity is substantially stronger for activism than for self-concept of ability. Socioeconomic status has the second strongest effect on activism (.198), whereas the effect of socioeconomic status is slightly stronger than ethnicity on self-concept of ability (.235). It may seem somewhat surprising

that sex has a negative effect on activism (–.123), while it has a positive effect on self-concept of ability (.115). These results imply that males are more active than females, while females have higher self-concepts of ability than males.

The coefficients reported in columns 3 and 4 illustrate that ethnicity has a fairly strong total causal effect (.194) on academic achievement. When sex, years of school, and socioeconomic status are controlled, non-Inuit students have, on average, an almost one-fifth standard deviation advantage over Inuit students. Nevertheless, the effect of ethnicity is reduced substantially when activism and self-concept of ability are included as intervening variables. Specifically, the effect of ethnicity is reduced from .194 to .053. This implies that approximately 73% of the effect of ethnicity is mediated by activism (68%) and self-concept of ability (5%).

Within this context, it is noteworthy that years of school and socioeconomic status also have relatively strong effects on academic achievement. The effect of years of school, as expected, is not substantially affected by the intervening variables. That is, students who have been in school for longer periods of time, irrespective of their social psychological dispositions, have higher scores on the academic achievement test than students who have spent less time in school. The effect of socioeconomic status, on the other hand, is substantially affected by the intervening variables. Specifically, the effect is reduced from .130 to .045 when the social psychological variables are included. This illustrates that approximately 65% of the total effect of socioeconomic status is mediated by activism and self-concept of ability. It is also noteworthy that the effect of sex is increased from .071 to .111 when the social psychological variables are included. Overall, this illustrates that females have higher academic achievement scores than males, but they would do even better if they had a more active disposition.

Finally, net of other variables, activism has a strong positive effect on academic achievement (.370). Specifically, a one-standard deviation change in activism results in a 37% standard deviation change in academic achievement. On the other hand, the effect of self-concept of ability is much smaller (.049). The importance of these two social psychological variables, particularly activism, is illustrated by the increase in the amount of variance explained in academic achievement. That is, when activism and self-concept of ability are added to the model, the explained variance is increased from 11.3% to 20.9%.

Discussion

In this study we have reported the effects of ethnicity on activism, self-concept of ability, and academic achievement. Moreover, the relationship between ethnicity and academic achievement has been assessed when activism and self-concept of ability are included as intervening variables and when sex, years in school, and socioeconomic status are included as independent variables. The results illustrate that Inuit students have less positive scores on activism and self-concept of ability than non-Inuit students. Moreover, ethnicity has a fairly strong total causal effect on academic achievement, with non-Inuit students obtaining higher scores

than Inuit students. The findings suggest that the effect of ethnicity on academic achievement is mediated in large part by activism and, to a much lesser degree, by self-concept of ability.

Finding that activism and self-concept of ability are important dispositions which mediate the effect of ethnicity on academic achievement may have practical implications. Specifically, these results suggest that it may be important for teachers of Inuit students to understand the social psychological dispositions of their students. That is, because Inuit students, in comparison with non-Inuit students, generally display lower activism in their classroom attitudes and lower self-concepts of ability, it may be important for teachers to encourage Inuit students to actively participate in the classroom. In this respect, it is suggested that a warm and demanding teacher may affect the Inuit students' self-concept of ability because the students become more active learners and experience greater academic success (Gecas, 1982).

More specifically, these findings provide support for the notion that certain teaching strategies may help improve the academic achievement of Inuit students by providing a better alignment between their existing dispositions and the methods used by their teachers. In this respect, Kleinfeld (1975, 1979) suggests that teachers of Inuit students require two attributes to be effective. First, effective teachers need to create an emotionally warm and personalized classroom environment. Second, effective teachers must demand that their Inuit students produce high quality academic work. In short, effective teachers of Inuit students must be "warm demanders" (Kleinfeld, 1975, p. 329). Such teachers focus on making Inuit students comfortable by personalizing their interaction with them, while at the same time demanding that they actively engage in their work and maintain high academic standards.

The link between activism and the recommendations that teachers become "warm demanders" is supported by recent research which suggests that the effects of family background and ethnicity on academic achievement is mediated, to a considerable degree, by social psychological variables (Cambell, 1983). In essence, this research literature suggests that for minority students a sense of control over their environment makes a significant contribution to their academic achievement. In reference to the items in our activism scale, it is reasonable to suggest that students who score low on this scale are more fatalistic, more present-oriented, and less prone to feel in control over their school experience. Difficulties emerge when students with such dispositions encounter a school system which relies on the value of future-orientation and control. In this respect, Bernstein (1961, p. 296) notes: "The school is an institution where every item in the present is finely linked to a distant future, and in consequence there is no serious clash of expectations between the school and the middle-class child."

However, when this standard model of school organization is provided for Inuit students who appear to be less concerned with control and future orientation, the misalignment occurs between the expectations of those who

manage the organization and their students. The responses of students to this type of strain may take a variety of forms. The typical response for Inuit students, however, appears to be the silent, passive resistance which is a coping strategy in line with the idea of fatalism and lack of control. Following the well-defined conditions for changing student conduct (Bredemeier & Bredemeier, 1978, pp. 167-187), Kleinfeld (1975) suggests that teachers provide genuine affective support that maintains Inuit student self-respect (the warmth component) while insisting that they participate in classroom learning activities (the demanding component).

Teachers may find it difficult to demand participation of Inuit students because they are generally quite passive (Kleinfeld, 1975, 1979). That is, Inuit students seem to have learned to use passive resistance as an effective coping mechanism and, "over the years [they have learned] that white teachers expect native students to stare mutely at the floor when confronted with an academic demand" (Kleinfeld, 1975, p. 327). Consequently, teachers who wish to be warm and demanding will likely need to exert considerable and consistent social pressure to resocialize their Inuit students. Clearly, this will involve a large expenditure of resources. At the same time, the fact that Inuit students are often silent and deferential to the authority of teachers makes it easier for teachers to direct their energies toward students who are doing well or to those who are disruptive. As a result, teachers who wish to develop a warm demanding relationship with Inuit students must be prepared to overcome considerable resistance associated with establishing new norms for both themselves and their students.

Although our findings are preliminary, the suggested implications are supported by more general social psychological research. For example, power and affectivity are two dimensions that appear related to significant changes in social psychological dispositions (Boldt, Lindquist, & Percival, 1976; Brim, 1966; Coser, 1979). This literature suggests that when substantial changes are desired, socializing agents need to have positive sentiments toward those they wish to help, while at the same time holding clear expectations for improved performance. The congruence between these dimensions and the warm and demanding dimensions used to interpret our findings is evident.

Furthermore, our suggestions are congruent with the profiles of the Inuit as reported in the culture and personality literature. This literature illustrates that ethnic groups vary in their susceptibility to the expectations for conforming with group norms (Boldt & Roberts, 1979). The Inuit have been studied in this regard, and the research supports what Briggs (1970) describes as their general "responsiveness to the needs of others" (p. 137). The existence of this disposition lends plausibility to the suggestion that Inuit students may respond to teachers who expect them to take an active orientation toward their school work if these expectations are reasonable and presented in a supportive manner.

Finally, even though we interpret our findings in terms of Kleinfeld's characterization of effective teachers for Inuit students, this is not the only framework that may be used to interpret these findings. The idea that effec-

tive teachers of Inuit student must be warm and demanding can be criticized on the grounds that it is an attempt to change the cultural dispositions of Inuit students. In other words, our suggestions infer that teachers should institutionalize a set of classroom practices that are not congruent with the cultural dispositions of Inuit students. In contrast, it may be argued that teachers should be developing strategies that build on rather than change the cultural dispositions of Inuit students. Such strategies may be more congruent with the idea of multiculturalism, but they would probably invoke radical changes in the existing cultural ideals and organizational arrangements of schools. From this perspective, teachers would need to change so that, for instance, they emphasize *being* rather than *doing*, *past and present* rather than *future*, and *harmony with nature* rather than *subjugation of nature*.

Although these types of changes in schools may be desirable, we do not know of any empirical research that describes and evaluates an effective educational system which has incorporated such changes. In fact, in Kleinfeld's (1975) research, two categories of teachers (sophisticates and sentimentalists) were reported as taking a more passive and understanding rather than active and demanding approach to their Inuit students. However, Kleinfeld's research also suggests that neither the sophisticates nor the sentimentalists were as effective with their Inuit students as those who were warm and demanding.

To argue that alternative teaching strategies have not been developed does not suggest that these strategies could not or should not be developed. Given the research that is available, conclusions about improving the education of Inuit students must be tentative. In this respect, our data and interpretations are suggestive rather than definitive. In essence, we conclude that Inuit education is in need of much more research if we are to develop more effective schools that would enable Inuit students to have more acceptable levels of academic achievement.

References

Alwin, D.F., & Hauser, R.M. (1975). The decomposition of effects in path analysis. *American Sociological Review, 40*, 37-47.

Barnes, G.E., & Vulcano, B.A. (1982). School self-acceptance among Canadian Indian, White and Métis children. *Canadian Journal of Behavioral Science, 14*, 60-70.

Bernstein, B. (1961). Social class and linguistic development: A theory of social learning. In A.H. Halsey, J. Floud, & C.A. Anderson (Eds.), *Education, economy, and society* (pp. 288-314). New York: Free Press.

Bienvenue, R.M. (1978). Self-evaluations of Native and Euro-Canadian students. *Canadian Ethnic Studies, 10*, 97-105.

Blishen, B.R., & McRoberts, H.A. (1976). A revised socioeconomic index for occupations in Canada. *Canadian Review of Sociology and Anthropology, 13*, 71-79.

Boldt, E.D., Lindquist, N.E., & Percival, G.A. (1976). The significance of significant others. *Canadian Review of Sociology and Anthropology, 13*, 345-351.

Boldt, E.D., & Roberts, L.W. (1979). Structural tightness and social conformity. *Journal of Cross-Cultural Psychology, 10*, 221-230.

Bredemeier, M.E., & Bredemeier, H.C. (1978). *Social forces in education*. Sherman Oaks, CA: Alfred Publishing.

Briggs, J. (1970). *Never in anger*. Cambridge, MA: Harvard University Press.

Brim, O.G. (1966). Socialization through the life cycle. In O.G. Brim & S. Wheeler (Eds.), *Socialization after childhood* (pp. 3-49). New York: Wiley.

Brookover, W., Beady, C., Flood, P., Schweitzer, J., & Wisenbaker, J. (1979). *School social systems and student achievement: Schools can make a difference*. New York: Praeger.

Cambell, R.T. (1983). Status attainment research: End of the beginning or beginning of the end? *Sociology of Education, 56*, 47-62.

Clifton, R.A. (1972). The social adjustment of Native students in a northern hostel. *Canadian Review of Sociology and Anthropology, 9*, 163-166.

Clifton, R.A. (1975). Self-concept and attitudes: A comparison of Canadian Indian and non-Indian students. *Canadian Review of Sociology and Anthropology, 12*, 577-584.

Clifton, R.A. (1977). Factors which affect the education of Canadian Indian students. In R.A. Carlton, L.A. Colley, & N.J. MacKinnon (Eds.), *Education, change, and society: A sociology of Canadian education* (pp. 183-203). Toronto: Gage.

Comber, L.C., & Keeves, J.P. (1973). *Science education in nineteen countries: An empirical study*. Stockholm: Almqvist & Wiksell.

Coser, R.L. (1979). *Training in ambiguity*. New York: Free Press.

Duncan, O.D. (1975). *Introduction to structural equation models*. New York: Academic Press.

Finley, D. (1983). Why Eskimo education isn't working. *Phi Delta Kappan, 64*, 580-581.

Franklyn, G.J. (1974). Alienation and achievement among Indian-Métis and non-Indians in the Mackenzie District of the Northwest Territories. *Alberta Journal of Educational Research, 20*, 157-169.

Fuchs, E., & Havighurst, R.J. (1973). *To live on this earth*. Garden City, NY: Doubleday.

Gecas, V. (1982). The self-concept. *Annual Review of Sociology, 8*, 1-33.

Hawthorn, H.B. (Ed.). (1967). *A survey of the contemporary Indians of Canada: Economic, political, educational needs and policies, Vol. 2*. Ottawa: Queen's Printer.

Heise, D.R. (1975). *Causal analysis*. New York: Wiley.

Honigmann, J.J., & Honigmann, I. (1970). *Arctic townsmen: Ethnic backgrounds and modernization*. Ottawa: Saint Paul University, Canadian Research Centre for Anthropology.

Inkeles, A., & Smith, D. (1977). *Becoming modern: Individual change in six developing countries*. Cambridge, MA: Harvard University Press.

Kahl, J.A. (1965). Some measurements of achievement motivation. *American Journal of Sociology, 70*, 669-681.

Kahl, J.A. (1968). *The measurement of modernism: A study of values in Brazil and Mexico*. Austin, TX: University of Texas Press.

King, A.R. (1967). *The school at Mopass: A problem of identity*. New York: Holt, Rinehart and Winston.

Kleinfeld, J.S. (1975). Effective teachers of Eskimo and Indian students. *School Review, 83*, 301-344.

Kleinfeld, J.S. (1979). *Eskimo school on the Andreafsky: A study of effective bicultural education*. New York: Praeger.

MacArthur, R.S. (1973). Some ability patterns: Central Eskimos and Nsenga Africans. *International Journal of Psychology, 8*, 239-247.

MacArthur, R.S. (1978). Ecology, culture, and cognitive development: Canadian Native youth. In L. Driedger (Ed.), *The Canadian ethnic mosaic: A quest for identity* (pp. 187-211). Toronto: McClelland and Stewart.

Marjoribanks, K. (1972). Ethnic and environmental influences on mental abilities. *American Journal of Sociology, 78*, 323-337.

Mueller, J.H., Schuessler, K.F., & Costner, H. (1977). *Statistical reasoning in sociology*. Boston: Houghton Mifflin.

Mueller, H.H, Mulcahy, R.F., Wilgosh, L., Watters, B., & Mancini, G.J. (1986). An analysis of WISC-R item responses with Canadian Inuit children. *Alberta Journal of Educational Research, 32*, 12-36.

Rampaul, W.E., Singh, M., & Didyk, J. (1984). The relationship between academic achievement, self-concept, creativity, and teacher expectations among Native children in a Northern Manitoba school. *Alberta Journal of Educational Research, 30*, 213-225.

Roberts, L.W., Clifton, R.A., & Wiseman, J. (in press). Exploring the value orientations of Inuit and White students: An empirical study. *Canadian Journal of Native Education*.

Taylor, L.J., & Skanes, G.R. (1976). Cognitive abilities of Inuit and White children from similar environments. *Canadian Journal of Behavioral Science, 8*, 1-8.

Thorndike, R.L. (1973). *Reading comprehension education in fifteen countries: An empirical study*. Stockholm: Almqvist & Wiksell.

Toohey, K. (1985a). English as a second language for Native Canadians. *Canadian Journal of Education, 10*, 273-293.

Toohey, K. (1985b). Northern Native Canadian language education. *Canadian Review of Sociology and Anthropology, 22*, 93-101

Vernon, P.E. (1966). Education and intellectual development among Canadian Indians and Eskimos. *Educational Review, 18*, 79-91, 186-195.

Vernon, P.E. (1984). Abilities and achievement of ethnic groups in Canada with special reference to Canadian Natives and Orientals. In R.J. Samuda, J.W. Berry, & M. Laferriere (Eds.), *Multiculturalism in Canada: Social and educational perspectives* (pp. 382-395). Toronto: Allyn & Bacon.

Wax, M.L., Wax, R.H., & Dumont, R.V. (1964). Formal education in an American Indian community. *Social Problems, 2*(4), supplement.

Wolcott, H.F. (1967). *A Kwakiutal village and school*. New York: Holt, Rinehart and Winston.

The Alberta Journal of Educational Research *Vol. XXXIV, No. 4, December 1988, 344-354*

GRACE V. MALICKY

and

CHARLES A. NORMAN
University of Alberta

Reading Processes Subgroups in a Clinical Population

The purpose of this study was to identify subgroups within a clinical reading population on the basis of how they processed information as they read. Oral reading miscues and retellings of stories were collected from 123 students in grades 1 to 9 who had been referred for diagnosis of reading problems. Results of factor analysis on miscue, retelling, and reading level variables revealed six processing factors which were consistent with an interactive theory of reading. A subsequent cluster analysis yielded six subgroups which differed in terms of how they processed information sources as they read. While there was some link between reading level and mode of processing, readers differed within, as well as across, reading levels, and these differences have pedagogical implications.

Much research on reading disabilities has focused on a search for pathology, or more specifically, on a single cause for a homogeneous phenomenon. From the disappointing results of this research, it is now widely accepted that children with reading problems form a heterogeneous group (Gittelman, 1985; Lipson & Wixson, 1986; Rutter, 1978; Singer, 1982).

Recent literature contains frequent calls for research on the identification of subgroups. Gittelman (1985) for example, notes that

> the identification of more homogeneous subgroups is a rational scientific as well as clinical goal, because better classification of the reading disorders should enable the selection of treatments tailored to each type of dysfunction. (p. 845)

Lipson and Wixson (1986) further state that research into differences and similarities within the reading disability group will help to understand the conflicting results of good-poor reader research and of intervention studies.

Initial attempts to identify subgroups of poor readers focused on underlying perceptual and cognitive variables as the basis for classification. However, Rutter (1978) indicates that caution is required in considering the implications of these classifications. Besides the fact that there is overlap across groups and children who do not fit into any group using these schemes (Rutter, 1978), research also shows that reading achievement is

---

Dr. Malicky is a member of the Department of Elementary Education and Director of the Reading and Language Centre.

Dr. Norman is with the Department of Educational Psychology specializing in the psychology and education of exceptional learners.

not improved by training the perceptual or psycholinguistic abilities involved (Gittelman, 1985; Glass, 1983). From these results, Rutter (1978) and Gittelman (1985) have advocated that differentiation of subgroups be based on measures which assess the reading process. However, few studies have moved in this direction. Lipson and Wixson (1986) report an unpublished study by Carr who used a profile of readers' performance on a battery of component skill measures to identify three distinct subgroups: balanced readers, active comprehenders, and phonological recoders.

From a theoretical perspective, the study reported in this article was based on the interactive theory of reading as postulated by Rumelhart (1977). He viewed reading as involving an interaction of different information sources available to readers, ranging from features of letters (bottom-up) through to syntactic and semantic knowledge (top-down). While good readers engage knowledge-based (top-down) and print-based (bottom-up) processing simultaneously, it has been hypothesized that poor readers may rely heavily on one type of processing and not enough on the other (Stanovich, 1980). The purpose of this study was to identify subgroups within a clinical reading population on the basis of how they processed information as they read.

Procedures

The sample consisted of 123 students who were assessed at the Reading and Language Centre, University of Alberta, between 1976 and 1987. There were 88 males and 35 females. Only subjects who were able to provide both oral reading miscues and unaided recall data on an informal reading inventory were included in the sample. All students were referred for assessment because of perceived reading problems, and at the time of assessment all were attending either regular classes in grades 1 through 9 or special education classes for children aged six through 15. Students were not excluded on the basis of IQ, and the sample included subjects from the mentally handicapped to superior range of intellectual ability. The average reading level of subjects in the sample was grade 3.3.

Passages from three commercial informal reading inventories were used to collect a sample of oral reading miscues and unaided retellings. The *Standard Reading Inventory* (McCracken, 1966) and *Classroom Reading Inventory* (Silvaroli, 1976) were used prior to 1984 to collect data from elementary and junior high age students respectively. Since 1984, the *Bader Reading and Language Inventory* (Bader, 1983) has been used. Both the *Standard Reading Inventory* and the *Bader Reading and Language Inventory* include an unaided retelling component in the test administration. The *Classroom Reading Inventory* was adapted to include this component as well. Each instrument consists of a series of word lists and graded passages at increasing levels of difficulty. Students began by identifying words on word lists, and the results on these lists were used to estimate the level at which to begin having students read passages. For each passage, subjects read the passage orally, recalled everything they were able to remember, and then answered questions to assess comprehension of information not included in the retelling. Passages at progressively more difficult levels were

administered until instructional and frustration levels had been obtained for each subject using criteria provided in the test manuals.

Oral reading miscues and unaided retellings were analyzed from passages at independent and instructional levels. Researchers have found that the nature of miscues and retellings at frustration level is significantly different than at independent and instructional level (Biemiller, 1979; Hood, 1982; Kavanagh, 1981).

Categories to analyze oral reading miscues were selected and adapted from Goodman and Burke's (1972) system. The categories, criteria used for coding, as well as the nature of information source indicated by miscues are noted in Table 1. Overall, miscues were viewed as providing information regarding how students were processing information *while* interacting with the text.

Unaided retellings provide another source of information on how readers process information, but retellings are collected *after* a selection is read, and hence reflect both reading and retrieval processes. Unaided retellings were divided into clausal units, and these were categorized and interpreted according to criteria presented in Table 2 (adapted from Fagan, 1984). In relation to Rumelhart's theory (1977), exact, specific, and embedded recalls reflect bottom-up or text-based processing, experiential recall reflects top-down or knowledge-based processing, and inferential recall involves interaction of both levels of processing.

Coding of the data resulted in five miscue and seven retelling variables. All raw scores were converted to percentages because of differences in the number of responses provided by subjects. These 12 processing variables, as well as the instructional reading level for each subject, were submitted to factor analysis. Six principal components with roots greater than one were rotated to the Varimax criterion. For theoretical reasons, the six-factor solution was selected to serve as the basis for cluster analysis (Veldman, 1967). The increase in error function was greatest between the six- and five-group solutions on the cluster analysis, and on this basis the six-group solution

Table 1
Categories, Criteria, and Interpretation of Miscues

Category	Criteria for Coding	Information Source
High Graphic Similarity	2 of 3 word parts in text word and miscue are the same	print
High Grammatical Acceptability	miscue is grammatical in relation to rest of passage	knowledge of language structure in relation to text
High Semantic Acceptability	miscue is meaningful in relation to rest of passage	knowledge of world in relation to text
Low Meaning Change	miscue results in no significant change to author's meaning	integration of knowledge and text
Correction	miscue is corrected	monitoring on basis of print, knowledge, and/or text

Table 2
Categories, Criteria, and Interpretation of Retellings

Category	Criteria for Coding	Information Source
Exact	verbatim recall of text	text
Specific	paraphrase of text	text from one clause
Embedded	combination of text units	text from more than one clause
Inferential	production of inferences	integrated use of knowledge and text
Experiential	elaborations not constrained by text	knowledge
Text Erroneous	erroneous recall of text	text
Knowledge Erroneous	erroneous use of background knowledge	knowledge

was selected for interpretation. Mean scores for all variables were calculated for each group.

Results and Discussion

Factor Analysis

The results of the six-factor solution are presented in Table 3. There was no clear achievement factor; instead, the achievement variable loaded on what appeared to be more obviously processing factors. Each of the six factors is interpreted below.

All the significant loadings on factor one were miscue variables and all three involved use of both knowledge-based and text-based information in an integrated manner. Hence this factor appeared to involve *integrative processing during reading*. The integrative nature of this factor was most evident in the meaning change variable because a reader must use both text information and background knowledge to produce miscues which are consistent with the meaning intended by the author. While grammatical and semantic acceptability appear to be more reflective of use of knowledge to predict meaningful and grammatical responses, miscues had to be consistent with the overall passage to meet the criteria set for these variables.

The second factor also appeared to be integrative, but this time it involved *integrative processing after reading* as indicated by the positive loading of inferential recall. Responses coded as inferential had to be consistent with the text; the reader basically filled in gaps in the text by integrating information from both knowledge and the text. It is interesting to note that this integrative factor, as opposed to factor one, was linked positively with level of reading, thus indicating that integration of information to make inferences was a function of reading proficiency. The presence of embedded recall on this factor is difficult to interpret because it appears that embedded recall involved a combination of text information only as opposed to the integration of both knowledge and text.

Table 3
Six-Factor Solution for Reading Level and Processing Variables

Variable	Factor Loadings 1	2	3	4	5	6
Instructional Reading Level	–0.055	0.536*	–0.497*	0.269	–0.219	0.285
Graphic Similarity	–0.140	0.037	0.779*	0.155	0.033	0.020
Grammatical Acceptability	0.893*	–0.037	0.083	0.066	–0.050	0.014
Semantic Acceptability	0.929*	–0.031	–0.047	–0.095	0.015	0.010
Meaning Change	0.692*	0.188	–0.412	0.059	–0.016	–0.029
Error Correction	0.131	0.098	0.287	0.387	0.659*	0.024
Exact Recall	–0.021	0.010	0.030	0.132	–0.117	–0.930*
Specific Recall	–0.008	–0.253	–0.602*	0.528*	0.095	–0.015
Embedded Recall	–0.024	–0.756*	0.027	0.305	–0.140	0.195
Inferential Recall	0.038	0.833*	0.165	0.069	–0.043	0.109
Experiential Recall	–0.028	0.107	0.378	–0.118	–0.534*	0.426*
Text Erroneous	–0.191	–0.005	–0.039	–0.326	0.749*	0.225
Knowledge Erroneous	0.008	0.027	–0.031	–0.790*	0.016	0.137

*Loadings greater than ±0.4

The achievement variable showed up again on factor three which appeared to be a *print-based processing* factor. The positive loading of graphic similarity coupled with the negative loading of instructional reading level suggested an inverse relationship between reading proficiency and reliance on graphic cues during reading. In other words, readers at lower levels of proficiency engaged in more print-based processing than those at higher levels. The negative loadings of meaning change and specific recall suggest that when readers focused heavily on graphic cues, it was difficult for them to attend to the author's meaning.

Factor four clearly involved *text-based processing* as indicated by the positive loading on specific recall. The reader must attend to and associate meaning with text in order to store and recall it after a selection has been read. The negative loading of erroneous knowledge-based information is also consistent with this interpretation.

Factor five was less clear-cut but appeared to involve *monitoring* in some way. The positive loading of error correction and erroneous recall appear to be contradictory because one would anticipate that correction of miscues would be related to more accurate rather than less accurate recall. These results may be an artifact of the way the error correction category was defined. Research tends to show that correction of miscues which are *not* meaningful is related to reading proficiency (Beebe, 1980; King, 1978). Because nonmeaningful and meaningful errors were combined in this study, it may be the correction of meaningful errors which is linked to inaccurate

recall of text. The other loading on this factor, a negative loading of experiential recall, revealed that when readers make corrections to achieve accuracy in word identification, they make minimal use of background knowledge to elaborate on what they read.

The sixth and final factor was again clear-cut and was labeled *knowledge-based processing* on the basis of the positive loading of experiential recall. This recall category involved elaboration on the text from background knowledge; information provided was not constrained by the text. The negative loading of exact recall served to strengthen the interpretation of this as a knowledge-based factor because exact recall was the most clearly text-based of all of the recall variables.

Cluster Analysis

The results of the six cluster solution are shown in Table 4 which includes means for each subgroup on each variable as well as means and standard deviations for subgroups combined.

A clinical approach was taken to the interpretation of these subgroups with differences of one standard deviation from the mean for the total group considered to have clinical significance. Differences of this magnitude are noted in Table 4. A clinical rather than statistical approach was taken to describe subgroups because it was felt that differences in patterns rather than on specific variables had the most potential for understanding how these children read. Subgroups are organized in Table 4 from those least able to integrate information (Subgroup 5) to the most integrated readers in this sample of clinic referrals (Subgroup 2). It is clear that increasing integration corresponds in only a general way to increasing reading proficiency.

More mean scores for Subgroup 5 were one standard deviation from total mean scores than were those of any other subgroup, and these deviations were evident on both miscue and recall variables. On miscues, students made less use of language and world knowledge to predict words but made more corrections than any other subgroup. On recall variables, they produced a high percentage of erroneous information and little verbatim recall. Generally, these students appeared to be confused about the nature of reading. They viewed accuracy of word identification as more important than comprehension of the author's meaning. These students were in grades 2, 3, and 7, or in special class. Hence, while most had only been in school for two or three years, one had been in school for much longer and was still confused about what reading is.

While at a lower instructional level in reading, Subgroup 6 obtained mean scores which were at least one standard deviation from the overall mean on three recall variables only. The high score on experiential recall indicated use of knowledge to elaborate on text information, but there was limited evidence of ability to constrain use of background knowledge as reflected in the relatively low score (nearly one standard deviation below the mean) on inferential recall. This subgroup also produced the most embedded recall units of any subgroup, revealing that when text was recalled,

Table 4

Mean Scores for Six Subgroups on Reading Level and Processing Variables

Variables	Means for Subgroups						Grand Means	Standard Deviation
	5 N=5	6 N-6	1 N=17	3 N=18	4 N=39	2 N=38		
Instructional Reading Level	2.56	1.66	3.57	3.71	2.84	3.85	3.31	
Graphic Similarity	35.80	49.17	24.82	8.83*	42.77	26.00	30.17	19.45
Grammatical Acceptability	36.80*	72.50	43.94	75.67	63.54	65.82	62.66	22.72
Semantic Acceptability	27.00*	56.67	25.94*	67.56	51.38	55.87	50.89	23.43
Meaning Change	27.00	22.67	25.06	68.83*	32.08	42.26	38.97	24.60
Error Correction	45.60*	20.83	5.24	19.94	32.51	15.34	21.56	18.53
Exact Recall	4.00*	4.00*	26.47	24.94	19.23	13.63	17.98	13.60
Specific Recall	9.00	16.33	31.00	32.28	19.87	15.40	21.23	12.63
Embedded Recall	4.00	31.17*	16.41	8.89	7.67	5.92	9.51	10.74
Inferential Recall	15.60	12.83	12.53	21.11	29.18	36.95	26.75	14.61
Experiential Recall	1.60	31.50*	3.71	2.44	5.54	9.40	7.13	10.30
Erroneous Text Recall	51.80*	0.00	7.41	7.00	8.80	7.47	9.25	11.66
Erroneous Knowledge	14.00*	4.17	2.18	2.72	2.44	9.68	5.24	8.54

* 1 standard deviation above or below the total mean

it was frequently combined. This may have been partly an artifact of the level these readers reached on the informal reading inventory; the sentences in early levels are generally much shorter than those which typically occur in the oral language of these students. Hence this may reflect use of language knowledge in recalling text. The pattern of responses on the miscue variables was different from that on retellings, with scores on all variables within one standard deviation of the means for the total group. Generally, this subgroup made use of both print-based information and knowledge to identify words when reading orally but was not yet able to integrate this information effectively on a comprehension task. The subjects in this subgroup ranged in grade levels from 1 to 4 with one student in a special class placement.

Although only one mean score for members of Subgroup 1 fell one standard deviation below the total means, mean scores on four other variables were either the highest or lowest for the total group. The score that was one standard deviation below the total mean was on semantic acceptability which assesses the reader's ability to rely on knowledge to predict words which make sense in relation to passage context. Subgroup 1 also obtained a relatively low score on inferential recall which again involves use of knowledge within the constraints of the text. Hence these readers made limited use of their background knowledge to predict words in passage context or to make inferences. They were distributed across all grades from 1 to 9 with many in the grades 2 to 4 range.

In contrast to members of Subgroup 1, those in Subgroup 3 made limited use of graphic cues to identify words. Despite the fact that members of this group processed fewer graphic cues than members of any other group, they were generally able to integrate the graphic cues they did analyze with their background knowledge to predict words which preserved the author's meaning. All scores on recall variables fell within one standard deviation of grand means, again suggesting that members of this subgroup were moving toward integration of information sources as they read for meaning. Students in this group were in grades 3 through 8 with most in the range of grades 4 to 6.

None of the mean scores of Subgroups 4 or 2 was one standard deviation away from the grand mean. However, there were some differences across these two groups on miscues, with Subgroup 4 relying more heavily on graphic cues coupled with a relatively high rate of error correction. Members of Subgroup 2 relied less heavily on graphic cues but were generally able to predict words which were consistent with passage meaning. Recall variables revealed that both subgroups were able to rely on text-based and knowledge-based information when reading for meaning, and both subgroups produced considerable inferential recall. Members of Subgroup 4 were fairly evenly spread across grades 2 to 6, and although they were fairly well-integrated readers, they did need to be somewhat less concerned with graphic cues. Their mean instructional reading level was somewhat below that of members in Subgroup 2 who ranged in grade placement from 1 to 9 with the majority again in grades 2 to 6. Students in group 2 appeared to be

well on their way to becoming effective readers. Some of these children were referred to the clinic to make sure they were making progress. Others were performing poorly in classrooms for reasons other than deficiencies in reading, or their teachers or parents may have overreacted to minor reading difficulties.

Summary and Implications

The factor analysis yielded six processing factors: one print-based, one text-based, one knowledge-based, two integrative, and one monitoring factor. These results fit well with Rumelhart's (1977) characterization of reading as involving an interaction of various sources of information with some factors being bottom-up (text-based), others more top-down (reader-based), and still others clearly involving an integration of information sources. The monitoring variable provides a regulative function and has been postulated by other theorists who work within the interactive framework (Brown, 1980; Yussen, Matthews, & Hiebert, 1982).

There was no clear-cut achievement factor; instead, reading level loaded on two of the processing factors. The negative loading of reading level on the print-based processing factor reflects an inverse relationship between reading proficiency and processing graphic cues. These results do not appear to be consistent with Stanovich's (1986) contention that

> It is not that good readers are less reliant on visual information, but that they expend less capacity to process visual information fully.... [Good readers] completely sample the visual array and use less resources to do so. (pp. 368-369)

While none of the subjects in this study could be called good readers, a higher level of reading achievement was associated with less graphic similarity of miscues, so it did appear that better readers were less reliant on visual information. Reading level loaded positively on the factor which involved integration of text-based and knowledge-based information to make inferences. This may partly reflect an instructional bias because poorer readers often receive more focus on literal than on inferential comprehension (Browne, 1971). While no causal links can be drawn from the results of the factor analysis, inclusion of work on drawing inferences would appear to be appropriate in the programs for most problem readers.

Results of the cluster analysis revealed that students in this clinical population varied widely in relative use and integration of information sources as they read. Generally, a higher instructional reading level was associated with a more balanced, integrated use of knowledge and text. However, at both lower and higher levels of reading proficiency, it was possible to identify subgroups on the basis of differences in how they were processing information. Because differences across groups appear to have pedagogical implications, the importance of going beyond a determination of reading level in the diagnosis of reading problems is reinforced by the results of this study. In order to demonstrate the nature of the pedagogical implications of this type of analysis, a few general program recommendations for different groups are presented below, recognizing that these are

far from complete and that further information would be needed in order to plan a program for any individual student.

A comparison of Subgroups 5 and 6 revealed the need for different emphasis in instructional programs despite the fact that members of both subgroups had difficulty integrating information sources. Members of Subgroup 5 demonstrated confusion on both miscues and recalls regarding the nature of reading and need to be immersed in meaningful reading activities. Members of Subgroup 6, while at a lower instructional reading level, were relatively successful in using both graphic cues and their own knowledge to predict words as they read but were far less successful at constraining use of knowledge by text on the recall task. They need to learn how to stay within the constraints of the author, and work on inferencing would help in this regard.

Even more demonstrative of the need for differentiation of instruction was a comparison of Subgroups 1 and 3, both at approximately a mid-third instructional reading level. Both subgroups included students across a wide range of grade placements. However, students in Subgroup 1 did not rely enough on background knowledge to predict words as they read, while students in Subgroup 3 did not make enough use of graphic cues. Both groups need to develop more effective strategies for identifying words, but Subgroup 1 needs to focus on using knowledge and Subgroup 3 on processing graphic cues.

Finally, the students in Subgroups 4 and 2 were already showing a fairly high degree of integration of information sources in their reading. Students in Subgroup 4 need a decreased focus on graphic cues, but generally both need to receive a well-balanced program with increased focus on reading to learn.

There is no intent in this article to imply that the six subgroups identified form a definitive classification of reading problems. Rather, the study reveals the feasibility of defining subtypes in terms of how children with reading problems process information as they read. Further studies using different measures and populations are needed to clarify the range of patterns of processing across problem readers.

References

Bader, L.A. (1983). *Bader reading and language inventory*. New York: Macmillan.

Beebe, M.J. (1980). The effect of different types of substitution miscues on reading. *Reading Research Quarterly, 15*, 324-336.

Biemiller, A. (1979). Changes in the use of graphic and contextual information as functions of passage difficulty and reading achievement level. *Journal of Reading Behavior, 11*, 307-318.

Brown, A.L. (1980). Metacognitive development and reading. In R.J. Spiro, B.C. Bruce, & W.F. Brewer (Eds.), *Theoretical issues in reading comprehension*. Hillsdale, NJ: Lawrence Erlbaum.

Browne, M.P. (1971). *An exploratory study of teacher-pupil verbal interaction in primary reading groups*. Unpublished doctoral dissertation, University of Alberta.

Fagan, W.T. (1984). *Categories for protocol analysis*. Unpublished paper, University of Alberta.

Gittelman, R. (1985). Controlled trials of remedial approaches to reading disability. *Journal of Child Psychology and Psychiatry, 26*(6), 843-846.

Glass, G.V. (1983). Effectiveness of special education. *Policy Studies Review, 2*(1), 65-78.

Goodman, Y.M., & Burke, C.L. (1972). *Reading miscue inventory: Procedures for diagnosis and evaluation*. New York: Macmillan.

Hood, J. (1982). The relationship of selected text variables to miscue scores of second graders. *Journal of Reading Behavior, 14*, 141-158.

Kavanagh, M.B. (1981). *Processing differences on three reading levels between no gain and gain groups in a remedial reading program*. Unpublished master's project, University of Alberta.

King, M.A. (1978). *An investigation of the differences in oral reading behavior between good and poor readers*. Unpublished master's thesis, University of Alberta.

Lipson, M.Y., & Wixson, K.K. (1986). Reading disability research: An interactionist perspective. *Review of Educational Research, 56*(1), 111-136.

McCracken, R.A. (1966). *Standard reading inventory*. Klamath Falls, OR: Klamath.

Rumelhart, D. (1977). *Introduction to human information processing theory*. New York: Wiley.

Rutter, M. (1978). Prevalence and types of dyslexia. In A.L. Benton & D. Pearl (Eds.), *Dyslexia: An appraisal of current knowledge*. New York: Oxford University Press.

Silvaroli, N.J. (1976). *Classroom reading inventory*. Dubuque, IA: Brown.

Singer, M.H. (1982). *Competent reader, disabled reader: Research and application*. Hillsdale, NJ: Lawrence Erlbaum.

Stanovich, K.E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, 16*(1), 32-71.

Stanovich, K. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*(4), 360-406.

Veldman, D.J. (1967). *FORTRAN programming for the behavioral sciences*. New York: Holt, Rinehart & Winston.

Yussen, S.R., Matthews, S.R., & Hiebert, E. (1982). Metacognitive aspects of reading. In W. Otto & S. White (Eds.), *Reading expository material*. New York: Academic Press.

The Alberta Journal of Educational Research *Vol. XXXIV, No. 4, December 1988, 355-374*

PHILIP NAGY

PETER J.A. EVANS

and

FLOYD G. ROBINSON
Ontario Institute for Studies in Education

Exploratory Analysis of Disagreement Among Holistic Essay Scores

This study, using a sample of 300 essays written on a descriptive and an argument topic, examined variation in holistic scores assigned by 20 experienced English teachers. In the first part of the analysis, focusing on the obtained scores, additional ratings were obtained for specific positive features, for errors, and for quality of argument. Results showed inconsistency of holistic judges across both topics and the range of essay quality. Degree of relationship between the holistic scores and the other types of ratings also varied across judges. In the second part of the analysis, holistic score distributions of essays eliciting disagreement were categorized; some characteristics of the writing could be linked to this categorization. Examination of the actual writing showed possible reasons for disagreement, but no one reason applied to more than a few essays. Although the focus was on classroom grading, the study has implications for large-scale assessment, including the areas of stimulus design, the range of acceptable responses, and scorer training.

Since the Britton, Martin, and Rosen (1966) and College Entrance Examination Board studies (Godshalk, Swineford, & Coffman, 1966), there have been many reports concerning the use of holistic (rapid impression) scoring of student writing. Some of these (Evans, 1981; Grobe, 1981) have been fine-grained analyses of possible reasons for the distribution of scores, but the issue of judge disagreement has not been examined in any systematic way. The present study specifically examines disagreement among judges and those essays which have produced this disagreement. First, patterns among judges and their inconsistencies across two topics and across the range of essay quality are examined. Second, the characteristics of the inconsistently scored essays themselves are investigated. These characteristics include error counts, specific positive features, and for one of the two

---

Dr. Nagy is Assistant Head of the OISE/UWO Educational Leadership Centre, the Ontario Institute for Studies in Education, London. His interests are in applied measurement and assessment of cognitive structures.

Dr. Evans is Head of the Ottawa Valley Centre of the Ontario Insitute for Studies in Education. His interests are in English language arts curriculum with special attention to evaluation.

Dr. Robinson is Head of the Midnorthern/Northeastern Centre of the Ontario Institute for Studies in Education, North Bay. His interests are in curriculum and instruction design from a cognitive science perspective.

topics, framework scores (Robinson & Stemp, 1986), in which the logic of the writer's argument is examined in detail.

Our focus is on grading at the teacher or classroom level. Those who provided the holistic scores used in this analysis—qualified and experienced high school English teachers—were untrained from the perspective of systematic large-scale assessment programs (Mullis, 1984). As is well known, such training procedures can greatly improve agreement, but in this situation can also interfere with the ability to generalize to the teacher working in the isolation of the typical classroom. Because we have uncovered some possible reasons why essay judges might disagree, our results are relevant to training for large-scale assessment. However, our main purpose is investigation of scoring at the school level.

Background

There is considerable literature on methods of scoring writing including that on holistic scoring, error counts, the assessment of positive (analytic) features, and the relationships between these methods. This study explores these systems and their interrelations in a context of holistic judge disagreement. The report makes use of the following terms:

1. holistic scoring—judges are asked to make a single, global quality judgment about each paper, reading rapidly for total impression (Evans, 1981);
2. judge—a holistic scorer. For purposes of clarity and consistency, only those providing holistic scores are referred to as judges, while those who provided other data are called raters, scorers, or (error) counters;
3. primary trait scoring—scoring is focused on the purpose of the piece of writing and success in attaining that purpose; detailed guides for scorers are devised to exemplify the criteria required for each point on a scale (Lloyd-Jones, 1977);
4. framework scoring—a variety of primary trait scoring in which a panel assigns ratings on several scales related to the quality of argument in persuasive writing (Robinson & Stemp, 1986);
5. positive feature scoring—also referred to in the literature as analytic scoring or positive characterization. Prominent single characteristics of writing are identified and each is rated according to quality (Mullis, 1984);
6. error counting—numbers of errors in several categories are counted according to clearly specified guidelines (Evans, 1981);
7. syntactic maturity—complexity of syntax is assessed by criteria such as number of words per T-unit, or number of words per clause (Grobe, 1981).

We need to distinguish between assessment of a particular piece of writing and assessment of writing ability. No claim is made that one or two pieces of writing obtained under stressful conditions will be a fair representation of ability. The present study concerns *pieces* of writing only and reaction to them.

Though holistic scoring had been employed earlier in numerous studies, the uses by Britton, Martin, and Rosen (1966) and Godshalk, Swineford, and Coffman (1966) appear to have been the springboard to its widespread use in many assessment projects (Evans, 1979, 1981; Gorman, White, Orchard, & Tate, 1982; McLean, 1982; Traub, Wolfe, Wolfe, Evans, & Russell, 1976). When particular sets of conditions, such as a large number of judges (Evans, 1981), careful monitoring with all judges on site (New Brunswick, 1977), and inclusion of common marker essays for calibration of judges (Evans, 1979, 1981; McLean, 1982; Traub et al., 1976) are met, reported reliabilities reach into the 0.80 to 0.90 range (Cooper, 1977, p. 19; Follman & Anderson, 1967). Charnay (1984) summarizes the conditions which ought to be present when holistic scoring is conducted. One of her chief concerns is the establishment of mode or topic, as will be discussed below in the context of probable causes of judge splits.

One problem with holistic scoring conducted without reference to explicit criteria is that it reveals no information on specific strengths and weaknesses; consequently, other methods have been pursued. For example, primary trait scoring (Lloyd-Jones, 1977) replaced holistic scoring in the American National Assessment. Variations of the primary trait approach have been applied across a variety of modes by Evans (1985, 1986) in studies in the Northwest Territories. Other large-scale studies of writing, in addition to using holistic scoring, have described results anecdotally with examples (Evans, 1979; McLean, 1982; Traub et al., 1976), or more systematically with reference to error frequencies, independent measures of syntactic maturity, or positive feature scoring.

There have been many pursuits of the question "What made the judges jump?" or "In what direction do judges tend to jump?" Most of the studies have been statistical: correlations, multiple correlations, and factor analyses (Evans, 1981). Studies of this type reach a number of similar findings, the most notable and frequent of which is that analytic scores or positive features correlate highly with holistic scores. The intent of the present study is to go beyond the reporting of relationships among scoring systems in the pursuit of reasons why these relationships exist. To this end, we have employed exploratory data analysis techniques (Tukey, 1977) capable of fine-grained scrutiny of the data.

Gorman et al. (1982, p. 98) report high correlations with holistic scores for scores on organization (0.77) and style (0.72), with grammar and spelling considerably lower. In Evans (1981), correlations with holistic scores tended to be at least as high across most topics and grades for both individual and combined positive feature ratings. McLean (1982) reports a correlation of holistic scores with aggregated positive features of 0.59 (p. 37).

In addition to positive features, length alone seems to have its rewards, as evidenced by correlations with holistic scores of 0.50 for number of words (McLean, 1982) and 0.60 to 0.70 for log (number of words, Evans, 1979). In an analysis of 24 potential predictors of holistic score, Grobe (1981) found that length was the largest unique contributor. Subsequent inclusion of

other variables in the regression equation suggested that *diction* also had a strong relationship to holistic scores. Evans reported that errors, length, and positive *characterization* (*features* in our terminology) accounted for 61% to 82% of the variance in holistic scores across 10 sets of essays. Positive features were the principal contributors in five of these sets, while length and error counts predominated in the remainder.

For errors, an originally exhaustive 19-category counting procedure (Evans, 1979) was later simplified by Evans (1981). These simpler categories, used here and in McLean (1981), are recommended because they capture most of the problems that cause difficulty for students and because they are easy to use. In this simplified system, errors are classified as (1) spelling, (2) major sentence errors including punctuation, and (3) grammatical errors, subcategorized as errors in (a) pronoun, (b) verb, and (c) other.

Several studies have examined the use of a number of positive features—explicit criteria—against which an essay is to be rated on a series of scales. For example, through factor analysis Diederich (1974) derived and elaborated four principal features: ideas, organization, wording (diction), and flavor (style), in addition to mechanics (usage, punctuation, spelling, handwriting). These criteria, as elaborated by teachers of grades 7 and 8 and edited by Evans, Brown, and Marsh (1977), are set out in the *Intermediate English Guideline* (1977) and the *Ontario Assessment Instrument Pool—Intermediate English* (1980). Adaptations involving criteria for organization, style, and language choice were employed in McLean (1982, p. 37) and adapted for use in the present analysis.

Framework retrieval scoring is an analysis of the logical elements and the relationships between elements in a student's essay (Robinson & Stemp, 1986). The main purpose of framework retrieval analysis is to yield a well-defined portrayal of an essay that can provide a more substantial focus in a follow-up writing conference than can spelling and grammatical errors. It is used in the present investigation to examine whether the logical framework produced is implicitly important in judges' holistic scoring.

Procedures

Essay and Judge Selection

The sample of essays used for the present study was obtained from an earlier study (Nagy & Jarchow, 1981), in which 320 essays on each of two topics were written by grade 10 students. The students were in academic, university-bound programs. About one half came from Newfoundland and the other half from Iowa. There were slightly more girls than boys, and the sample was a bit more rural than the North American average. Thus schools tended to be small, with between one and four classes of academic grade 10 students.

The topics were a description, "Describe trying to get to sleep with a mosquito buzzing" (Topic 1), and an argument, "Argue for or against the statement: 'Children over fourteen should have as much say as parents in reaching decisions which affect the entire family'" (Topic 2). Brand and

House (1987) have investigated differential performance and rating of quality on emotionally loaded and neutral topics and have found relationships between essay rating and emotional level (by self-rating) of the writers. Gee (1987) has found that essay scores in choice situations are related to topic choice.

In the original study, each essay was holistically scored by four experienced grade 10 English teachers. Two variables, a quality score and a measure of interjudge agreement, were produced for each essay after norming within the 40 scores awarded by each judge. These scores were used in the present study to draw samples of 150 essays within each topic:

1. Fifty essays which elicited the most disagreement among the four judges on each topic became the inconsistent essays.
2. From the remaining essays, on which there was relative agreement among the four judges, a random sample of 50 from the 20th to 40th percentile in each topic became the low consistent essays.
3. A random sample of 50 from the 60th to 80th percentile in each topic became the high consistent essays.

Once the selection of the sample from the larger set was completed, no further use of the Nagy and Jarchow (1981) results was made. Each group of 50 essays was divided in half to produce two sets of 75 essays within each topic, 25 from each of the three categories: inconsistent, low consistent, and high consistent. The three types of essays were intermingled for presentation to judges, and several orders of presentation were used. There was no direct control of the actual order of scoring.

The judges in the present study were all qualified, experienced grade 10 English teachers who either volunteered for participation as a result of previous involvement with the investigators or who were nominated by their department heads.

Scoring Procedures[1]

Holistic scores. As our focus was on typical classroom practice, judges were not specifically trained in systematic procedures. Each participant was asked to judge rapidly on a scale of 1 (low) to 10 (high) without applying explicit criteria, to ensure that all points on the scale were used, and to check himself or herself if essays started to *pile up* on only one or two points on the scale. Forty judges participated, each grading one set of essays on each topic, for a total of 150 essays. Twenty holistic scores for each essay were produced.

Error counts. Detailed instructions were prepared as to what constituted an error. Errors were grouped as spelling, grammar, and sentence structure. Eight counters, none from among the holistic judges, counted errors in one set of essays each, thus yielding two independent error counts for each essay. The original intention was to average the two counts. However, there was so much disagreement on the error counts that a third counter was added, and the middle of the three counts used.

Positive features. Four positive features—organization, viewpoint, language, and structure—were rated on separate 1-5 scales following detailed instructions. Viewpoint was applied to Topic 1 only, as this rating was considered redundant in view of the framework scoring used for Topic 2. Each essay was rated by two independent scorers on three or four features, depending on topic. The ratings on each feature were averaged, except in cases of extreme disagreement when the opinion of a third referee was used.

Framework retrieval. Framework scoring (Robinson & Stemp, 1986) is appropriate only to argument or position essays, and thus was applied only to Topic 2. Two teachers, working closely with one of the principal investigators, rated the essays on the following six criteria, using scales from 0 to 3: alternatives considered, type of criteria (consequences) considered, evidence for claims, logical flow, structure for decision making, structure for process argument. Ratings were obtained through discussion by searching for a dominant position on each criterion in each essay rather than by looking for consistency in the position taken.

Data Analysis

The overall purpose of the study was to track judge disagreement by identifying the criteria implicitly used in the holistic scoring process. There were two distinct aspects to the data analysis. The first, essentially quantitative and reported immediately below, focused on the judges rather than the essays and dealt entirely with the measures—holistic scores, positive features, error counts, and framework retrieval scores—rather than with the actual writing. As the objects of this analysis were the holistic scores, the perspective taken was that all other data collected were *correct*. Except for the framework scoring, which was applied only to Topic 2, the two topics are reported together, facilitating a comparison across the two.

The analysis proceeds largely using exploratory methods developed by Tukey (1977). The purpose of this method, as the name suggests, is to *explore* relationships within the data set, rather than to *test* statistical hypotheses. Our aim in the research was to examine possible reasons for disagreement, rather than to report sizes of correlations between various measures. Since Tukey's original work, acceptance of his ideas by the orthodox statistical community can be attested to by publication in the Wiley Series on Probability and Mathematical Statistics of Hoaglin, Mosteller, and Tukey (1983).

The analysis began with a simple cross-tabulation of median scores awarded by each judge to the two topics, designed to explore inconsistencies in generosity at a global level. Next, a more detailed analysis was undertaken to explore consistency across the range of essay quality. Each of the four sets of essays was broken into five "quintiles" of 15 according to overall quality (mean of the 20 scores). A two-step median polish (Tukey, 1977) was performed on the scores, resulting in a set of residual deviation scores, independent of the median generosity of each judge and the median quality of each essay.

These values were then summed for each judge within each quintile, producing five values indicative of consistency across the quality range. These values yielded an assessment of judge consistency both relative to other judges and for the same individual across the two topics. We allowed a tolerance of 5% before characterizing a judge as either positively or negatively inconsistent within each quintile.

These complex data were further reduced, yielding a rating of high, moderate, or low consistency for each judge across the quality range and across topics. For ratings across the quality range, the absolute difference between number of positive inconsistent ratings in the top two versus the bottom two quintiles was added to the corresponding value for negative inconsistent ratings. Counts of 0 or 1 were considered highly consistent, 2 and 3 moderate, and 4 or more low. For the rating across topics, a count was made of the number of quintiles for which the ratings (positive, negative, or tolerable) were the same. Counts of 5 were highly consistent, 3 or 4 moderate, and 0, 1, or 2 low. Individual judges' score distributions were used to exemplify the patterns of inconsistency uncovered by this analysis.

Following this analysis, essays were categorized as high, low, or inconsistently scored. These characterizations were based on the scoring for this study only, and are independent of the similar categorization done in the original Nagy and Jarchow (1981) study. All entries in the residual matrices from the median polish were converted to absolute values, and essay (row) totals were summed. The 25 essays with the highest such totals from each set were designated as *inconsistently* scored. The remaining 50 from each set were divided into 25 *high* and 25 *low* scored. Patterns within the positive features, error counts, and framework scores for each of these groups were examined using stem-and-leaf diagrams. Patterns among these three scoring systems were investigated using scatter-plots.

Despite some caution with respect to the use of the correlation statistic with skewed data, correlations were calculated between each judge's holistic scores and each of total error counts and total positive features. These correlations were divided into high, moderate, and low using cutoffs of 0.40 and 0.50 for relationships with the positive features, and 0.30 and 0.40 for those with the errors. The correlations were then cross-tabulated. This type of analysis was also used to investigate whether judges' implicit reliance on error counts, positive features, or framework scores was consistent across the range of essay quality.

Finally, for the examination of the actual writing, a more fine-grained categorization of scoring distributions assigned to each essay was required. Because of its post hoc nature, this categorization is detailed in the results section below. The second analysis, essentially qualitative, focused on the essays rather than the judges and examined in detail a subset of essays scored inconsistently. Although the numerical values are discussed, the actual writing was the target of examination.

Results of Score Analysis

Overall Generosity in Holistic Scores

Table 1 contains the distribution, in cross-tabular form, of the median holistic scores awarded by each of the 40 judges on the two topics. A median quality essay was scored anywhere between 30% and 70% on the 10-point scale depending on the judge and topic. Individual judges were not internally consistent; only 18 of 40 gave the same median score for both topics. For example, the 23 judges who awarded a median score of 5 for topic 2 gave the following median scores for Topic 1: one 3, five 4s, ten 5s, and seven 6s. Even allowing that quality on the two topics might be different, this lack of consistency is striking. Similar results showing a relationship between topic choice and score were obtained by Gee (1987).

Consistency Across Quality Ranges and Topics

In attempting to find reasons for scores varying as they do, it is important to ascertain whether judges are consistently generous across the quality range, or whether, for example, some are unusually generous with poorer essays. The quintile technique described above was used to assist in finding different patterns within the scores awarded by each judge. Table 2 summarizes the judges as they were placed in consistency categories by this analysis. It is evident that our ad hoc criteria for these categorizations were more stringent across topics than across the quality range. However, more than one half the judges were rated low on one or the other criteria, and three were rated low on both.

This technique, by providing substantial data reduction and allowing emergence of patterns in the scores, allowed easier interpretation of raw score distributions for each judge. Examples, mostly from Topic 1, are contained in Table 3. The quintile summation showed Judge 28 as harsh on good quality essays and easy on poorer ones; from Table 3, 60 of his or her 75 scores were between 4 and 7. From the quintile summations, Judge 1 was hard on the better quality essays; from Table 3, Judge 1 gave 50 of 75 scores of 4 or less. Other judges' patterns in Table 3 are those of Judge 33, who packed 60 of 75 scores between 5 and 8, and Judge 34, who was one of

Table 1

Crosstabulation of Median Scores of Judges Across Topics[1]

	Topic 1 Median			
Topic 2 Median	3	4	5	6
4		1	5	
5	1	5	10	7
6			3	7
7				1

[1]Entries are numbers of judges.

Table 2
Judge Consistency Summary[1]

Consistency Across Score Range	Consistency Across Topics		
	High	Moderate	Low
High	3	5	10
Moderate	4	5	7
Low	2	1	3

[1]Entries are numbers of judges.

the few judges perfectly consistent across all five quintiles within our 5% tolerance. The last two columns of Table 3 show the distribution for Judge 7, who was one of the judges showing the greatest difference across the two topics. For Topic 1, Judge 7 awarded 14 scores of 8 or more, while for Topic 2, only one.

After discussion of patterns within the positive features, error counts, and framework retrieval scores, we return to the holistic scores and their relationships with other essay characteristics.

Patterns in Positive Features

The essays in each set were classified as high, low, or inconsistently scored as described above.

The pattern[2] of positive features for the inconsistently scored essays was closer to that of the high scored essays than the low scored ones. The most separation between high and low scored essays was in Organization and Sequence, and Sentence Structure. There was less for Language Choice, and the least for Point of View. This trend was reflected across topics. Also of interest is the fact that there was considerable overlap in the positive features of the high and low scored groups. Typically, only about 20% of each group (low or high) were rated well or poorly enough in each feature to be beyond the range of the other group in that feature (e.g., overlap of the distributions was about 60%). This suggests that by and large the aspects of writing assessed by the independent positive features were not of paramount importance in the overall holistic scores.

Patterns in Error Counts

Correlations among error counts for the three counters were in the 0.40 to 0.70 range for each of spelling, sentences, and grammar, as well as the total count (calculations based on Topic 1, Set 1 only). While such values do not inspire great confidence in the error counts of an individual rater, the choice of the middle of three values for each count ensures that the error counts were reasonably reliable and valid. However, given the apparent difficulty in reaching agreement in error counting, it is legitimate to ask whether the holistic judges, who did not participate in the error counting, were sensitive

Table 3
Frequency Distribution of Scores of Selected Judges

	Judge Number					
Score	28	1	33	34	7	7[1]
1	1	8		6	2	2
2	2	10		5	4	7
3	6	21	3	14	4	9
4	10	11	7	6	15	15
5	17	7	14	11	12	16
6	16	6	16	9	14	13
7	17	6	18	9	10	11
8	2	2	12	8	6	
9	2	2	1	5	7	1
10	2	2	4	3	1	

[1]Refers to Topic 2; all others are Topic 1.

to the error frequencies and types to the same degree as were the independent error counters.

Several characteristics of the error count data are striking. First, the data were positively skewed; most essays had few errors, while a few had many. Second, the low scored and high scored essay error distributions overlapped considerably; low scored essays tended, on balance, to have more errors. Third, spelling errors were the most common type. Virtually all the high scored essays had relatively few spelling errors, but the essays with many spelling errors divided about evenly between low scored and inconsistently scored. This suggests that one of the main sources of disagreement among judges is the importance they place on errors, especially spelling errors. This interpretation is supported by a fourth observation, that essays with remarkably few errors also elicited disagreement. That is, freedom from errors satisfied some judges, but not others. Finally, grammatical errors seemed to have the least relationship with holistic scores. These data, especially those concerning spelling and grammatical areas, receive strong support from Evans' 1981 study.

Patterns in Framework Retrieval Scores

Framework retrieval was rated on six criteria using scales from 0 to 3. Stem-and-leaf diagrams revealed that the traits for which points were awarded were relatively rare; results for most subscales were quite positively skewed. There was at best moderate separation between high and low scored essays on most scales, the major exception being Logical Flow. Where any pattern could be discerned (Consequences, Logical Flow, for example), inconsistently scored essays tended to similarity with the high rather than the low scored essays. As an overall comment, holistic judges did not seem to be actively searching for the attributes which these scales demand in order to award high or even moderate holistic scores. This suggests that

these qualities are in general outside the normal expectations of the holistic judges.

Patterns Across Scoring Systems

Relationships between error counts and positive features, as well as among the error counts as examined through scatter plots, tended to be negligible. There was no evidence that inconsistent holistic scores were related to disparities either between error counts and positive features, or among the different error counts. Relationships of the framework scores with the positive features tended to be moderately positive, and with the error counts close to zero.

Of major interest is the relationship of holistic scores to the other values obtained. We deal with this issue in two analyses: the first, immediately below, deals with differences across judges; the second, in the preamble to the qualitative analysis in the next section, focuses on the essays in a detailed examination of characteristics which elicited disagreement.

Table 4 summarizes the relationships of judges' holistic scores to other criteria for the two topics, based on the correlational analysis and categorization described above. As can be seen, there is considerable variation in what may be the judges' implicit criteria for scoring. One trend evident from Table 4 is that more judges' scores were highly related to positive features for Topic 2 than for Topic 1. Not evident from Table 4, but also of interest, is that only one of 40 judges appeared in the corresponding box for both topics, while 18 appear in boxes that do not "touch." Apparently, judges do not maintain their implicit criteria across topics.

A second type of consistency was sought by examining the relationship of an individual judge's scores to the external criteria across different qualities of essay. Using an analysis similar to that used to produce Table 4, it was

Table 4

Crosstabulation of Strength of Relationship of Holistic Scores to Error Counts and Positive Features[1]

Strength of Relationship to Error Counts		Strength of Relationship to Positive Features		
		High	Moderate	Low
	Topic 1			
High		6	2	6
Moderate		6	2	3
Low		8	7	0
	Topic 2			
High		13	2	0
Moderate		9	4	2
Low		7	0	3

[1]Entries are number of judges.

found that judges implicitly gave positive features more weight in better quality essays in about 15% of the cases (essay-judge combinations) and gave positive features more weight in poorer quality essays in another 20%. In about 15% of the cases, judges gave more weight to errors in better quality essays (all in Topic 2), and in another 15% of the cases they gave more weight to errors in poorer quality essays (almost all in Topic 1). About one half the judges showed inconsistency in the implicit criteria adhered to as quality changed, and about one half showed different patterns of behavior on this issue across the two topics.

Correlations of each judge's scores with the total of the framework retrieval scores ranged from 0.08 to 0.72. There were noteworthy patterns. First, of the five judges with correlations over 0.60, three had the lowest relationship with both positive features and error counts. This strongly suggests that these three judges were marking implicitly from within a framework retrieval perspective, a perspective different than that of all others in the sample. Second, of the three judges whose correlations with framework scores were below 0.20, all three had strong relationships with error counts, and two also had strong relationships with positive features. These findings, suggesting identification of criteria implicitly used by holistic judges, are offered as a prelude to the analysis involving the actual writing.

Qualitative Examination of the Essays

Preliminary Comments

To this point in the analysis, the data surrounding the essays have been examined, but the writing itself has not. The next step in the analysis was conducted for each topic on the 50 essays found to produce most scorer disagreement. Before proceeding, we present, with the benefit of hindsight, some comments on the wording of the stimulus topics.

Topic 1 was descriptive: "Describe trying to get to sleep with a mosquito buzzing," a topic which one can assume is within the experience of all students. The phrase "Describe trying to get to sleep," however, led many students who were attempting to write on a universal level—what *one* does or feels—either to produce almost abstract writing or to confound their viewpoint, switching between *you* and the first person narrative. The variety of responses to the stimulus is one likely source of judge disagreement.

Topic 2 was an argument: "Argue for or against 'Children over fourteen should have as much say as their parents in reaching decisions which affect their entire family'," a topic obviously within the experience range of grade 10 students and on which personal experience can be drawn. There were problems as well with this stimulus: Students were directed to argue an extreme position rather than to present a balanced, compromise view. The phrase "as much say as" was extreme and not clear, and the scope of "decisions" was open to interpretation. Students were left to make too many decisions concerning how best to proceed; thus, although the stimulus was common in one sense, it produced a wide range of not particularly comparable responses. Thus, while for Topic 1 it can be assumed that judges are comparing fundamentally the same kind of writing in making judgments,

this assumption cannot be made for Topic 2. Some of the variability in holistic scores can be attributed to this problem.

Essay Classification

For the remainder of the analysis, a more elaborate scheme was required for categorization of essay score distributions, rather than the simple split into high consistent, low consistent, and inconsistent. For each essay the frequency distribution of holistic scores was examined. A check was made of outlier scores to ascertain whether any judge had been scoring erratically enough to require elimination; there seemed to be no case for deleting judges. Each frequency distribution was examined and classified into one of four categories (example score distributions are given in Table 5):

1. Acceptable: acceptable central tendency, with a few stray scores, or distributed in a concentrated rectangle in the upper or lower range, probably indicative of masking of essential agreement through implicit use of a 1-8 or 3-10 scale (1135 and 2143);
2. Rectangular: fairly even distribution across many points with no apparent mode (1050);
3. Modal Rectangular: like Rectangular but with a number of scores concentrated at one point (1046) or two or three adjacent points (2077);
4. Bimodal: two modes some distance apart (1038), one mode and one concentrated rectangle (2055), or tending to trimodal (1131).

The final three categories, Rectangular, Modal Rectangular, and Bimodal were designated as the inconsistently scored essays for further analysis of the actual writing. Of the original 50 categorized as inconsistently scored in the first part of the analysis, there were 42 from Topic 1, the description, and 41 from Topic 2, the argument.

Analysis of Other Scores

Ratings on each positive feature, error counts, and framework retrieval scores (Topic 2 only) were matched against holistic scoring patterns. For the positive features, only 12 of 42 descriptive and two of 41 argument essays had ratings on individual positive features differing from any other by more than 1 on a 5-point scale. Thus the summed rating for the features was used as the measure of positive features. For the error counts, the total count (per 100 words) was employed. In the essays under examination, error frequency was greater than five in only 10 instances (nine descriptive and one argument), and in every case but two the loading was on spelling. For the framework retrieval, the total score was taken as a measure of quality of argument.

A hypothesis to explain lack of judge agreement is that split scores tend to occur in essays rated high on positive features but containing frequent errors. That is, some judges tend to pay more attention to or give more weight to errors than others. For convenience, an error frequency of 4 was taken as a cutoff for *problem* errors, and summed ratings of 12 (of a possible 20) for Topic 1, and 10 (of a possible 15) for Topic 2 as the dividing line between high and low positive features. For the 30 high error essays in our subsets,

Table 5
Sample Score Distributions for Selected Essays

Scale	Distribution Category								
	Acceptable		Rectangular	Modal Rectangular		Bimodal		
	1135[1]	2143	1050	1046	2077	1038	2055	1131
1		4						4
2		3	1					2
3		4	2	2	1	1	2	2
4		4	2		1	1	2	3
5		5	3	3	2	4	3	4
6	2		3	6	2	6	5	1
7	6		3	3	5	1	1	4
8	4		3	4	4	4	6	
9	4		3	2	2	3	1	
10	4				3			

[1]Essay identification numbers

20 descriptive and 10 argument, there were 12 descriptive essays and only one argument essay with high positive feature ratings. This may indicate some tension between positive features and errors in description but not in argument. However, judge splits occurred with considerable frequency even when both error counts were high and positive features low, so this tension was not a major factor in disagreement.

Many studies have shown a fairly strong correlation between word count and score obtained. Overall in our total sample, about a third of the descriptions and a fifth of the arguments exceeded 225 words. In the *disagreement* subsets about half the descriptions and one quarter of the arguments were similarly long. Thus, for exposition at least, no case can be made that length was a special contributor to the unusual score patterns; for description, it may have contributed to a limited extent. Within the descriptions, there were four instances of high length/low positive features, three of them in the Bimodal category. For description, few essays in the set under study were *short*; there were more for argument, but in most cases brevity was associated with a low positive features rating. Hence a brevity versus positive feature rating tension is virtually absent and does not appear to play a significant role in judge disagreement.

The positive features and framework retrieval scores are two different ways of assessing quality. Using cutoff values of 10 in both cases, the two ratings of quality agreed for seven of eight cases in the Rectangular group, five of 11 Modal Rectangular, and only seven of 19 for the Bimodal essays (Topic 2 only). It appears that some judges have been inattentive to the logic of argument, as there was little relationship between features of organization and style and the quality of argument when it was rated independently. This, then, may be a major factor, especially in the bimodal essays where agreement was found in only one third of the cases. It is

reasonable, therefore, to hypothesize that the tension between effectiveness of style and logic is a source of judge disagreement.

The Descriptive Writing—Topic 1

Rectangular essays. The problem of voice (universal or personal) and viewpoint described above was apparent in the 11 rectangular essays. Four writers became directly trapped in switches between *I* and the universal case, and two others came close. Four produced consistently good first-person narratives, and one solved the problem by turning the description into a humorous *scientific* study. The net effect was that judges were required to compare essays actually written in somewhat different modes.

In several instances, the writers started out formally with some finely crafted sentences and later deteriorated into a colloquial style, characterized by frequent *ands* or *buts* and even sentence fragments in an attempt at narrative immediacy. A device frequently used in these essays was hyperbole: Some were quite witty, and some fell flat. Hyperbole is an obvious source of humor, but unevenness in execution combined with differing tastes was probably a contributor to variations in grading.

Modal rectangular essays. The difference between this group of essays and the rectangular group was quite striking. Most of these essays could, with some qualification, be judged relatively easily along one major dimension with one major deficiency over which some judges might be likely to split. The existence of a mode suggests that many judges, in noting defects as well as strengths, *took an average* rather than coming down strongly on one side of the fence or the other.

Of the 12 essays in this group, two appeared to be wholly indifferent pieces: little to praise and little to condemn. A third, a reasonably good narrative, was marred by problems suggestive of English as a second language. Another was a first-person narrative of a mental hospital patient attempting some dialogue with the mosquito—a delightful, light touch, but incoherent and careless in its switches—imagination excellent, execution doubtful. The remaining eight were characteristically dull.

Bimodal essays. Logic would suggest that the bimodal distribution would be produced when more judges were more strongly committed to (or against) a particular feature of the writing and less willing to average strengths and weaknesses than was evident in the modal rectangular distribution. Examination of the group did indicate that something like that tended to happen. Two essays, which were well written, were wholly or almost wholly off topic. Half the judges put one essay in the 5 range; half gave it 1. In the other essay forgiveness at the higher end was greater: two 10s, a 9, and an 8 for style; five judges gave the compromise 5 or 6; and there were seven 1s, three 2s, and a 3.

Three essays were judged weak with judges split between 3 or 4 against 1. Does one award *something* for a piece reasonably on topic and complete despite its errors, or does one sum *undistinguished* and *errors* and come out close to zero?

Another essay produced its major split between 10-9 and 7, but two judges scored it at 2. It was a first-person narrative, telling the story in the past, and had many positive features. Diction, on the whole, was good with some imaginatively hyperbolic expression. However, one also found "blue-assed mosquito," "a damn mosquito," and some version of *damn* repeated six times in the piece. It would seem that while most judges appreciated the essay, the judges who gave it 7 discounted the language choice, and those who gave 2 were possibly so annoyed that they *damned* the whole thing.

Several essays illustrate the problem previously noted—the incomparability of essays quite differently constructed owing in part to a lack of clarity in the stimulus. The range of scores is strong evidence of the difficulty judges had in coping with widely different interpretations and treatments of the stimulus. One essay, with a bimodal distribution of scores grouping around 7-8 and 3, was a refreshingly humorous but quite short exposition on "how to get to sleep." Another, lengthy and with adequate detail, was indifferently told; judges split around 6, 4, and 2. A third was quite long, but there were numerous clinkers in grammar and diction which apparently infuriated some judges as scores gathered around 7, 5, and 1. Finally, one essay produced a range of scores from 9 down to 1 with the concentration at 7-6 and 4. Here the student evidently decided to take seriously and literally the problem of describing how *a person* might get to sleep. The description was carefully planned with a strong concluding sentence. Thus a well-organized essay, complete and relatively error-free, nevertheless produced a wide array of scores, probably because it was unexpected, and, barring the conclusion, intrinsically uninteresting.

The Argument Writing—Topic 2

Problems with the stimulus. Because of the extreme form of the stimulus, with *for* or *against* directing the student to take an immoderate position rather than to present a balanced, compromise view, there were problems for many students. Most students stuck to argument, however, and only one shifted into narrative. Although students might have drawn on personal experiences, few did, and only one of the 41 shifted away from the topic to any extent. On the whole, the essays examined here were abstract and dull.

In an informal observation to one of the authors, Britton (Britton et al., 1966) noted that in his study judges tended to split over essays in which the student expressed an unpopular viewpoint. This suggestion led to the hypothesis tested here. Twenty-three of 41 split-score essays argued fairly unambiguously "Yes, 14-year-olds should have an equal voice in reaching decisions," but two thirds moderated that position to "in most cases." It is therefore impossible to make any general claim that splits tended to occur in response to an *unpopular* stance. However, in several instances where mechanics and quality of argument were both dubious, the student appears to have found redemption in the eyes of some judges by expressing the right sentiments about teenagers and parental wisdom. Thus one may be

tempted to say that judges tended to bunch when appropriate sentiments were expressed.

It is important to note that a large majority of students, whether they argued for or against, qualified their position when in effect the wording of the topic prohibited such a moderate approach. Some students who did stick strictly to the unreasonable extreme prescribed may have hurt their scores (there are too few cases to generalize), and others certainly found themselves out on limbs from which they had to climb down.

Language, mechanics, and organization. The inconsistently scored essays did not seem to suffer in any systematic way from difficulties in language, mechanics, or organization, although there were problems across a substantial number of essays. In each subset there were a number of adequately organized and argued essays which had serious mechanical deficiencies: rectangular, three of eight; modal rectangular, four of 11; bimodal, eight of 22. Several were colloquial or drowned in clichés. Some judges came down heavily against the poor quality of language while others did not. At the other stylistic extreme, a number of the rectangular and modal rectangular essays exhibited an unusual awkwardness, possibly due to efforts to work out of the extreme positions initially chosen.

Length may also have been at issue in a number of the splits and disagreements: The trend was to short, even single paragraph, responses. A few long but otherwise undistinguished essays did receive a number of high as well as mid-range scores, while short essays tended to receive mid- to low-range scores. This trend was not unique to any set but characteristic of all.

Logical frameworks. A separate investigation was made of the Framework Retrieval scores against splits in holistic scoring. Of the 41 essays, 17 obtained a score of 12 or greater on quality of argument. The highest framework score was obtained by an essay for which holistic scores provided a rectangular distribution across 6-9 (also two 10s, a 5, a 4, and a 1!). It was long and was the one essay that shifted somewhat off topic, carrying the argument and illustration back to age six. It did, however, exhibit logic and provided examples. Here, evidently, other factors besides quality of argument caused the score disagreement.

Three essays received the next highest quality of argument score. All three had modal rectangular holistic score distributions, and modes at 7. One of the three was an awkward *the person* pro and con evasion of either position. Another was long, working through an extended illustration, with numerous flaws. The third, though exhibiting vigorous style, ended on a flat note, and the route to any conclusion was difficult to discern. Thus, for each of the three, a mixture of pluses and minuses along with variable quality of argument probably contributed to judge disagreement.

Conclusions

Summary

This investigation proceeded along two paths: a quantitative analysis of holistic, positive features, error, and framework scores, and a qualitative

analysis of the actual writing in the essays which elicited holistic judge disagreement.

The quantitative analysis demonstrated patterns of inconsistency among judges, both across and within 40 individual judges and two essay topics. Because of the large number of possible stimulus types available, generalization from one argument and one description to *writing* should be done with caution. About 25% of the judges scored Topic 2 harder than Topic 1, and another 25% did the opposite. Only three judges were found to be highly consistent across both quality range and topics; other judges avoided high scores, low scores, both high and low scores, and in one case, high scores for one topic and low for the other. The distribution of positive features of the inconsistently scored essays was more like that of the high than the low scored ones. Essays with remarkably few errors tended to elicit judge disagreement on holistic scores, showing that, for some judges, absence of error was quite important, while for others it was not.

In the analysis of the actual writing, disagreement among holistic judges could not be attributed to large differences among the scores assigned for the different positive features. Numerous compositions where judges split could be characterized as *the good try, inconsistent success phenomenon*. Within the inconsistently scored descriptive essays, several subsets could be identified with various features: error counts and positive features at odds, problems of voice, or one dominant feature over which judges might disagree. For the inconsistently scored argument essays, there was a subset which exhibited disharmony between writing style and quality of argument, and another characterized by good logic and mechanical problems. However, the central problem was once again the stimulus itself. Many students got themselves trapped in an effort to support an extreme position and had difficulty in extricating themselves. Efforts to do so often led to an unfortunate style.

Implications

In the qualitative analysis, few clear, general trends emerged. Nevertheless, a number of probable sources of judge tension were uncovered and deserve particular attention in the training of judges in large-scale assessment. Attention to these features may also be of value to the classroom teacher in less formal writing assessment. In position writing there seems to be a premium on writing that expresses *appropriate* sentiments. If there is a not-so-well hidden agenda here concerning the criteria that *really count*, that agenda needs to be brought into the open and fully discussed so that judges will be conscious of the factors to be balanced. In descriptive writing several types of treatment are possible in response to most stimuli. Judge training should include attention to the variety of approaches students employ so that legitimate alternatives are recognized and credited.

The register selected (or shifted to) clearly is an aspect of writing style likely to cause judge disagreement, especially concerning humor, efforts at hyperbole, and stretching the boundaries of *good taste* in diction, imagery, and even usage. It is important to examine these styles in order to make

some consistent accommodation to them. Judges who expect formal writing may be critical of informality of style, even when effectively done.

The stimuli in this study have been stumbling blocks for the students and the judges. Besides care in design, already discussed, pretesting the stimulus is also necessary to minimize an unexpected variety of response modes, a problem noted by Stratta and Dixon (1982). Another much longer study by Ruth and Keech (1982) examined the complex issue of topics or *prompts* generally. They argue that testers "should consider methods of cuing which avoid sounding too much like directions to the writer and more like genuine options. Cues should be arranged if possible to suggest a natural way of planning the response to the topic" (p. 148). Those setting writing tasks must give careful thought as to what really is being asked of the student; it is not acceptable that *best guessers* get the rewards.

In ongoing work related to this project (details are available from the authors), one of the authors is currently pursuing the hypothesis that holistic judges might be especially afflicted by the limits on human information processing, in that they have about a minute to read up to a page of text and react to it. Thus judges may dwell on relatively superficial aspects not because they regard these as most important, but because they cannot do sufficient processing in the time allowed to progress to deeper levels of processing. In preliminary analysis, think-aloud protocols from eight judges were used to identify a rough four-point scale of depth of processing: (a) rudimentary matching of stimulus to a stored image, experienced judges respond automatically to simple mismatches; (b) periodic assessment of how sensible the text being processed *sounds*; (c) relating two components in the stimulus such as noting lack of subject-predicate agreement; (d) mapping stimulus elements to cognitive structure to assess whether they are meaningful or sensible. Such work may yield further insights into appropriate procedures for scoring essays, both for large-scale assessments and classroom grading.

Acknowledgments

We wish to thank Roslyn Klaiman for extensive assistance in reducing the bulk of the original report.

This research was funded by a grant from the Ontario Institute for Studies in Education.

Notes

1. Because of space limitations, details of the scoring procedures have been omitted. Further information is available from the authors.
2. Because of the amount of data, the stem-and-leaf diagrams used for this analysis of positive features, error counts, and framework scores cannot be included. These tables are available from the authors.

References

Brand, A.G., & House, G. (1987). Relationships between types of assignments, writer variables, and emotions of college composition students. *Alberta Journal of Educational Research, 33*, 21-32.

Britton, J.N., Martin, N.C., & Rosen, H. (1966). *Multiple marking of English compositions. Examinations Bulletin #12, The Schools Council*. London: HMSO.

Charnay, D. (1984). The validity of using holistic scoring to evaluate writing—A critical overview. *Research in the Teaching of English, 18*, 65-81.

Cooper, C.R. (1977). Holistic evaluation of writing. In C.R. Cooper & L. Odell (Eds.), *Evaluating writing: Describing, measuring, judging* (pp. 3-31). Urbana, IL: National Council of Teachers of English.

Diederich, P.B. (1974). *Measuring growth in English*. Urbana, IL: National Council of Teachers of English.

Evans, P. (1979). *Evaluation of writing in Ontario: Grades 8, 12 and 13. Review and Evaluation Bulletin, Vol. 1, #2*. Toronto: Ontario Ministry of Education.

Evans, P. (1981). *The analysis of writing assessment: The OISE writing evaluation project 1977-1980*. Ottawa: Ontario Institute for Studies in Education.

Evans, P. (1985). *Northwest Territories middle years achievement testing program 1985—Writing*. Ottawa: Ontario Institute for Studies in Education.

Evans, P. (1986). *Northwest Territories middle years achievement testing program 1986—Writing and reading*. Ottawa: Ontario Institute for Studies in Education.

Evans, P., Brown, P., & Marsh, M. (1977). *Criteria for the evaluation of student writing, grades 7 and 8—a handbook*. Toronto: OISE Press.

Follman, J.C., & Anderson, J.A. (1967). An investigation of the reliability of five procedures for grading English themes. *Research in the Teaching of English, 1,* 190-200.

Gee, T.W. (1987). Drafting and revising processes in grade 12 students' examination writing. *Alberta Journal of Educational Research, 33,* 96-114.

Godshalk, F.I., Swineford, F., & Coffman, W.E. (1966). *The measurement of writing ability*. New York: College Entrance Examination Board.

Gorman, T.P., White, J., Orchard, L., & Tate, A. (1982). *Language performance in schools, secondary survey report #1, assessment of performance unit, for the Department of Education and Science*. London: HMSO.

Grobe, C. (1981). Syntactic maturity, mechanics, and vocabulary as predictors of quality ratings. *Research in the Teaching of English, 15,* 75-86.

Hoaglin, D.C., Mosteller, F., & Tukey, J.W. (1983). *Understanding robust and exploratory data analysis*. New York: Wiley.

Intermediate English guideline. (1977). Toronto: Ontario Ministry of Education.

Lloyd-Jones, R. (1977) Primary trait scoring. In C.R. Cooper & L. Odell (Eds.), *Evaluating writing: Describing, measuring, judging*. Urbana, IL: National Council of Teachers of English.

McLean, L.D. (1982). *Report of the 1981 field trials in English and mathematics, intermediate division*. Toronto: Ontario Ministry of Education.

Mullis, I. (1984). Scoring direct writing assessments: What are the alternatives? *Educational Measurement Issues and Practice, 3*(1), 16-18.

Nagy, G.P., & Jarchow, E.M. (1981) *Estimating variance components of essay ratings in a complex design*. Presented at the Annual Meeting of the American Educational Research Association, Los Angeles.

New Brunswick Department of Education. (1977). *Writing assessment reports*. Fredericton: Author.

Ontario assessment instrument pool—intermediate English. (1980). Toronto: Ontario Ministry of Education.

Robinson, F., & Stemp, H. (1986). A framework approach for improving student writing. *Indirections, 2*(2), 45-54.

Ruth, L., & Keech, C. (1982). Designing prompts for holistic writing assessments: Knowledge from theory, research and practice. In L. Ruth (Ed.), *Properties of writing tasks: A study of alternative procedures for holistic writing assessment* (pp. 32-213). Berkeley, CA: National Institutes of Education.

Stratta, L., & Dixon, J. (1982). *Teaching and assessing argument. Discussion Booklet 2*. Southampton, UK: Southern Regional Examinations Boards.

Traub, R., Wolfe, R., Wolfe, C., Evans, P., & Russell, H.H. (1976). *Secondary-postsecondary interface project II: Nature of students* (2 volumes). Toronto: Ontario Ministry of Education and Ministry of Colleges and Universities.

Tukey, J.W. (1977). *Exploratory data analysis*. Don Mills, ON: Addison-Wesley.

The Alberta Journal of Educational Research *Vol. XXXIV, No. 4, December 1988, 375-389*

JOE FRIS

and

JAMES H. BALDERSON
University of Alberta

Leaders' Priorities and the Congruity Imperative

Contemporary systems theorists and researchers often refer to the congruity imperative: the requirement that organizational activities, procedures, and structures remain congruent with the exigencies of the environment. However, in the research on the behavior of school administrators there is little direct and conclusive information about the extent to which superintendents and principals monitor the external environment and adjust priorities accordingly. This study examined the work priorities of reputedly effective leaders of Canadian school organizations for evidence of the degree to which they were driven by the congruity imperative. The data provided a demographic profile of those effective leaders, a typology of their most important administrative issues, and a typology of the imperatives underlying those issues. Content analysis of the reported imperatives provided support for the idea that the work priorities of effective leaders of school organizations are partly determined by the congruity imperative.

The long-term viability of organizations is linked to the *fit* between the exigencies of the environment and the organizational arrangements (structures) that regulate activities in an organization (Burns & Stalker, 1961; Katz & Kahn, 1966; Lawrence & Lorsch, 1967; Mintzberg, 1973; Thompson, 1967; Tosi & Carroll, 1976; Woodward, 1965). According to contingency theory, one of the essential functions of administrators is to ensure that organizational activities, procedures, and structures remain congruent with the demands of the environment. Tosi and Carroll (1976) have expressed this congruity imperative:

> All organizations are in an active relationship with their environments. Just as any organism, such as [the human being], must learn to adapt to its environment or perish, so too must organizations.... [Hence] managers will find it necessary to reassess continually the relationship of the organization's structure and goals to the values of the various publics, internal and external, with which it interfaces. (pp. 155-156)

It seems the most common orientation to the congruity principle is the here-and-now perspective; however, some authors (Aguilar, 1967; Morrison,

---

Dr. Fris is a member of the Department of Educational Administration and editor of *The Canadian Administrator.*

Dr. Balderson is also with the Department of Educational Administration and is the publisher and editor of the *Canadian School Executive.*

Renfro, & Boucher, 1983) point to the need to take an anticipatory approach. For example, Renfro and Morrison (1983) postulate that

> We tend to spend our time on issues we perceive to be most important right now; we fail to scan our surroundings for changes that are in the early stages of development. The flood of problems that forces us into crisis management makes concern for emerging issues appear to be a luxury. It is not. It is a necessity. (p. 50)

In regard to the matter of finding ways to update structures and procedures as circumstances change, Mintzberg (1973) found that senior administrators consider it important to take full charge of the organization's strategy-making system. The main reason for this is that "as [information] centre, the [senior] manager can best ensure that significant decisions reflect current knowledge and organization values" (p. 77). In a similar vein, Peters and Waterman (1982) observed that effective executives "are dependent on lots of people, not just a few in central R & D" (p. 286). The added implication, then, is that managers are likely to be more effective if they also consult people outside their own organizations for solutions to their problems and challenges.

In sum, the literature on effective organizations indicates that executives must (a) systematically monitor the environment for changes that may necessitate adjustments in how the organization functions, (b) ensure that practices in the organization reflect current knowledge and are congruent with relevant conditions in the organization's environment, and (c) involve others in the process of identifying significant environmental pressures. In addition, the literature indicates that the congruity imperative is particularly important when the environment is volatile and that it is of particular significance to senior administrators because of their boundary spanning role. However, studies of the work performed by school superintendents and principals provide few insights on the extent to which they are influenced by the external environment.

For superintendents and principals, the importance of staying abreast of changes in society is accentuated by the fact that education is essentially a social service. Further, superintendents and principals operate at the boundaries of their organizations. These facts suggest that school system and school effectiveness are linked to the degree to which superintendents and principals satisfy the congruity imperative by responding to the exigencies of their organizations' external environments.

However, there is little direct information on the extent to which school superintendents and principals consciously pay attention to the congruity imperative (Morris, 1984) and the information we do have is frequently indirect and ambiguous (Crowson, 1987). We do not know with any certainty whether superintendents and principals systematically track changes in the external environment and deliberately initiate activities that are seen to be necessary to preserve or improve the organization-environment *fit*. Nor do we have strong evidence on the relationship between the attention that administrators pay to the congruity imperative and the effectiveness of their educational units. The information that is available tends to be enigmatic, if not contradictory (Crowson, 1987; Payne, 1987): Some of the research sug-

gests that superintendents and principals tend to have an inward focus, some indicates a sensitivity and responsiveness to forces in the external environment.

Crowson's (1987) review of literature on the superintendent's role revealed many interesting contradictions and anomalies, among them gaps and inconsistencies in data on the superintendent's control over the functioning of his or her system. For example, Hannaway and Sproull (1978-79) observed that coordination and control activities by central office staff have relatively little effect on what happens in schools, least of all on curriculum and instruction. Further, Crowson's review of Mintzberg type studies indicated that much of the large interpersonal dimension of the superintendent's work is devoted to *choreographing* rather than directing others' activities (Cuban, 1985; Feilders, 1978; Larson, Bussom, & Vicars, 1981; Peterson, 1984; Pitner, 1978; Pitner & Ogawa 1981). Yet, as Larson et al. (1981) noted, most superintendents focus on the internal management of their school districts. Contrary to an image which is widely accepted, superintendents spend relatively little time interacting with school trustees and boards; most of their time is devoted to interactions with subordinates.

Similarly, McLeod's (1984) study of directors and superintendents in Ontario revealed a tendency to focus on the internal management of the system. His data reflected "[a] preoccupation with the organization's orderly functioning, as well as the prevention of nasty disruption to the normal course of the educational enterprise" (p. 179). The data also indicated an inclination to avoid the transformational leader role—working at the cutting edge of change and progress in the profession. Instead, McLeod concluded that the CEOs in his sample tended to adopt the *functionary* role—assuring that the most rational means for achieving socially desirable ends are implemented: "The [CEO] is drawn into activities where credibility as pioneer of educational change and progress is likely to be subordinated to mundane strategies designed to keep matters under control" (p. 188). McLeod suggests that this happens because of a tendency to evaluate an administrator's performance "in terms of what does *not* go wrong" (p. 180, emphasis added), with the result that control and technical proficiency take precedence over ventures that require initiative, creativity, and independent action.

Irrespective of the nature of the leadership role or style adopted, senior administrators spend a lot of time interacting with others. Friesen and Duignan (1980) found that approximately 70% of the total working time of the superintendents they studied was spent in verbal contacts with individuals or groups of individuals and that those contacts were distributed as follows: 57% with *internal* persons/groups—students, teachers, school administrators, central office staff; 23% with *boundary* persons/groups—trustees; 9% with *external* persons/groups; and 10% with unclassified others. Moreover, almost 26% of the *contact* time was devoted to obtaining information. It is important to note, however, that Friesen and Duignan's data provide no information on the extent to which the superintendents' con-

tacts, whether with *internals* or *externals*, provided information about developments in the external environment; nor do we know whether any of the time spent on planning strategies for system operation (19%) was prompted by the congruity imperative or by *maintenance* concerns.

Payne's study (1987) was prompted by the contradiction between the conventional wisdom that parents are denied any significant influence on the school and the research findings which suggested that the school principal might in fact be significantly "constrained and channeled by parental influence." Payne sought to determine whether "parents act as a control mechanism over principal behavior and whether parental influence is encouraged by superordinates in order to constrain principal behavior." Payne concluded that

1. Principals were sensitive to the concerns of parents and tried to ensure that parents were satisfied.
2. Principals, in order to maintain a balance among the various influences on them, used parents as "monitors of various environments for the school."
3. Principals' responsiveness to pressures from parents varied: They tended to be more responsive in matters of general policy, less so when specific technical core and personnel matters were at issue.
4. Superordinates expected principals to "contain parent bodies' tendencies to exceed the role envisaged for them."
5. Principals at times increased their capacity to exert influence on superorinates by organizing and encouraging parents.
6. Superordinates used parents to influence principals' behavior.
7. Principals perceived parental influence as a constraint fostered by the district.
8. The *loosely coupled* conceptualization of school systems is misleading: "Schools are highly penetrated. Principals, to maintain the legitimacy of the school, have to cope with diverse ... constituencies ... and particularly they must satisfy the parent community." (pp. 323-326)

Clearly, in the school system that Payne studied, principals were sensitive to the exigencies of the environment.

Two sets of findings by Grainger (1984) provide inferential data on the manner in which principals deal with environmental demands. First, he found a discrepancy between the levels at which parents desired and achieved participation in curriculum, instruction, and personnel decisions. Most parents wanted to upgrade their role from the advisory/support level to the sharing level. However, the majority of educators in Grainger's study felt that such decisions required professional expertise that the parents did not possess and therefore preferred to exclude parents from this domain. Moreover, the parents reported that they were denied the organizational structures that would have permitted them their desired decision making role. These findings suggest that the educators may have protected themselves against direct pressures from the environment by deliberate manipulation of the organizational framework, that is, by using the buffering technique.

Second, Grainger's data indicate that parental attempts to influence affairs in the school system had not been denied absolutely but were merely channeled:

> Principals referred to the considerable influence exerted on them as a consequence of the countless informal contacts they had with parents when opinions and concerns were expressed.... There was common agreement [among principals and teachers] ... that school personnel sought to accommodate what they defined as basic community values and beliefs in their administration of the school. (p. 327)

This suggests that principals may act as *sensors* at the boundaries of their organizations but only as part of a *short-range warning system*, one that is tuned in to only the immediate community.

Principals, like superintendents, also seem predisposed toward a functionary role. Gunn (1984), after examining the job satisfaction of high school principals, concluded that the image of the principal as a strong instructional leader was not supported by his investigation:

> Many respondents saw themselves as being in a "middle management" position rather than in a top leadership position as head of an institution. Their perceived leadership role might be described as an "institutional maintenance" role in which the primary function was to maintain the status quo or to "keep the ship afloat." (p. 233)

As evidenced by the recent studies reviewed here, research on the activities of school superintendents and principals has provided little direct or conclusive information about the extent to which they monitor the external environment and adjust priorities accordingly. Indeed, there are puzzling anomalies: Some of the research provides evidence of active monitoring and responsiveness to the external environment; other research indicates that principals and superintendents focus primarily on smoothing the internal workings of their organizations—on *maintenance* rather than *adaptive* functions (Tosi & Carroll, 1976). Yet other research indicates a *gatekeeping* response to environmental pressures. Furthermore, there are intriguing indications that the locus of the monitoring and adaptive functions may be outside the ranks of educators and that superintendents and principals may tend to take a lower level *functionary* role in adapting their organizations to changed environmental conditions.

To clarify the effect that the congruity imperative has on the behavior of school administrators, it would be helpful to have data on the following questions:

1. What are the issues or problems that principals and superintendents have given the highest priority?
2. Why are those issues given high priority?
3. To what extent are those priorities influenced by environmental pressures?

In this article we report findings from an exploratory investigation of these questions. More specifically, the objectives of this exploratory study were to generate information on:

1. the specific administrative issues that a select group of Canadian superintendents and principals had given the highest priority;
2. the types of issues that were being addressed and the relative frequency with which those types were designated *most important*;

3. the imperatives that caused the issues to be given the highest priority; and
4. the loci of those imperatives—internal to the organization or external.

Method

Chief educational officers, superintendents, and directors of school jurisdictions and senior administrators of selected educational organizations in Canada ($n=768$) were asked to nominate the three most effective educational administrators known to them. The response rate for this nominations phase of the investigation was 24%.

Each of the nominated persons ($n=219$; many were nominated more than once by different nominators) was then sent a questionnaire. Ninety-two completed and returned the questionnaire. Only one respondent elected to remain anonymous by obliterating identifying information on the questionnaire. Follow-up interviews were conducted by telephone with 26 of the respondents; on average these lasted 45 minutes.

In the initial section of the questionnaire, the respondent was asked to record personal demographics—sex, office in the organization, highest academic qualification, and experience (administrative and other) in education.

The remainder of the questionnaire called for unstructured responses. To identify the most pressing issues that confront school executives, respondents were asked to provide a random listing of significant current and anticipated administrative problems/issues and a ranking of those problems/issues in terms of personal priority. Then the respondents were instructed to focus on only the issue that had been ranked most important and to provide the following information: (a) a detailed description of the issue, (b) reasons for giving it the highest priority, (c) a description of the strategies being applied, (d) a description of supportive factors, (e) a description of hindering factors, (f) an estimate of resolvability, (g) an estimate of the time required to resolve the issue, and (h) other pertinent comments or advice.

Content analysis, coupled with multirater validation, was used to reduce the 752 problems/issues that were listed by the principals and superintendents in the sample to 18 categories.

Responses to the question "What are the reasons for giving this issue/challenge a high priority?" were first classified, then cross-tabulated by office/role to ascertain whether principals and superintendents were influenced by different types of imperatives.

Results

The respondents were stratified by position as follows: principals/vice-principals, 25%; superintendents/directors, 54%; others (provincial/federal/academics), 21%. The latter were excluded from the analyses reported here.

Highest formal education levels were Bachelor's, 2%; Bachelor's+, 12%; Master's, 27%; Master's+, 36%; Doctorate, 21%; Other, 2%.

Experience in education ranged from 11 to 40 years, with a mean of 26 years. Experience in educational administration ranged from five to 35 years, with a mean of 18 years.

Types of Issues

The request to "take a few moments to bring to mind the problems/issues that you have already encountered and those which you expect to come up against in the near future" produced a total of 752 entries—many of them duplicates or variants. Content analysis revealed that they could be grouped into 18 classifications:

1. Collective bargaining,
2. Curriculum,
3. Declining enrollments,
4. Evaluation,
5. Financial,
6. French,
7. Governance,
8. Legal,
9. Managerial,
10. Organizational,
11. Planning,
12. Plant and property,
13. Policy,
14. Program,
15. School community relations,
16. Student personnel,
17. Teacher personnel,
18. Technology.

Most Important Issues

The next question addressed was "Which types of issues were most often designated very important?" The hierarchy of types presented in Table 1 emerged.

Clearly, the three classes of issues that were of critical concern to the principals and superintendents in this sample related to the management of teaching personnel, finances, and program. Less frequently identified, but still of the highest priority among these school executives, were problems associated with governance, management, and student personnel. The remaining high priority issues were identified by just a few of the respondents.

On the surface, this ordering suggests that administrators were more frequently concerned about teachers than they were about their students;

Table 1

Hierarchy of Types of Issues

Type of Issue	Percentage Ranking Issue Most Important
1. Teaching personnel	23%
2. Financial	16
3. Program	15
4. Governance	8
5. Managerial	8
6. Student personnel	7

however, as will become evident later, the data provided clear evidence that concern for students' welfare was at the root of practically all issues. As one director of education insisted, "The bottom line is to ensure that students learn; that is why we are all here!"

Principals' and Superintendents' Priorities

When the priorities of principals and superintendents were analyzed separately, some interesting parallels and contrasts emerged. The superintendents' priorities were as follows: (1) teaching personnel, (2) finances, (3) program, and (4) governance. The teacher personnel class of issues was given the highest priority far more frequently than the other three classes. The principals' priorities were as follows: (1) teaching personnel, (2) student personnel, (3) management, and (4) finances. The highest priority was given far more frequently to the first two categories than the other two.

Nature of the Issues

Teaching personnel issues. Ranked *most important* by 23% of the sample, these issues related to supervision and evaluation, reduction of staff, reassignment and recruitment, welfare and morale, tenure, and relationships with unions. Supervision/evaluation and welfare/morale issues were reported most often.

Three themes regarding supervision and evaluation were apparent. The strongest and most consistently articulated theme was a commitment to improving service to clients—students and society. This concern mirrored the conviction that student achievements and other outcome measures are linked in the first place to the quality of teachers' performance. The second theme was a concern about teachers' resistance to supervision and evaluation, particularly in districts which have older staffs. This was frequently linked to a reluctance to change; sometimes it was attributed to an apprehension that "the motives for formal policy on teacher supervision are other than those ... stated." There were frequent references to the need to work with teachers to allay such fears. The third theme was articulated most explicitly by a New Brunswick superintendent who wrote: "Supervision is a massive problem because [every] teacher, supervisor, and administrator is involved. Consequently the issue cannot be solved with any kind of hit and miss action. The solution ... is to have a Master Plan."

Concerns about welfare and morale, although most directly associated with teachers, also reflected an underlying concern for the welfare of students. A principal noted that

> *the importance of teacher morale cannot be underestimated [because] teachers are of prime importance in the educational enterprise ... [yet] the provincial Ministry has attacked teachers ... as a group of undedicated people who are not doing their jobs ... with the result that teachers feel unappreciated both for their work in the classroom and for their extra-curricular responsibilities.*

As for other causes of morale problems, some respondents noted that the role of a teacher today is much more complex than it was in the past and that lack of public understanding of this complexity has led to a frustration and demoralization of the profession. Other respondents attributed the

morale problem to declining enrollments and the attendant erosion of job security. Yet others referred to teacher welfare problems such as stress-related medical problems, resistance to new ideas and change, and the stagnation that occurs when an older and static staff is not exposed to the more recently developed knowledge acquired by younger teachers.

In sum, the challenges for this group of respondents were to develop professional competence, counteract apathy, expose staff to new ideas, increase morale, cope with stress, and promote healthy lifestyles. The overarching goal was to "improve the quality of service to students ... and the community."

Financial issues. Ranked *most important* by 16% of the respondents, the main concern here was the effect of budget cuts on the delivery of services to students.

In the words of one superintendent, "the crux of the issue is how to trim budgets, reduce programs, and lay off staff ... without doing permanent damage." A common aspect of this concern was exemplified in the comment that "without appropriate funding and program flexibility, we will not be able to accommodate a wide range of student needs and interests." A related concern was the possibility that "the erosion of services will create a gradual flight from public school education to private schools."

The fiscal challenges identified by this group of school administrators also translated into the following specific concerns: the struggle to manage reductions in fiscal resources, a task made even more difficult by new courses and programs expected by parents and mandated by provincial legislation; the maintenance and renovation of older buildings; the reduction of administrative staff "without losing the support and leadership of some of these people."

Only one school executive articulated a positive point of view regarding financial constraints. While agreeing that the financial resources of the district were strained and that the elimination of programs was quite possible, he insisted that "this new set of circumstances makes the game more challenging and exciting!"

Program issues. This type of issue was ranked *most important* by 15% of the respondents and reflected three general concerns.

First, a number of school administrators identified specific programs that they wished to introduce, improve, or reorganize. Their goals included such things as expansion of extension programs, introduction of a human ecology program, development of diagnostic testing programs, delivery of special education services, and so on.

Second, numerous respondents were wrestling with the implementation of provincially mandated renewal projects. Those projects related to programs for kindergarten through to secondary students, reviews of secondary school credit systems and compulsory courses, school organization, and the effectiveness of delivery services in general.

Third, some administrators were concentrating their efforts on dealing with pressures to promote excellence. Some noted the difficulties generated by disparate conceptions of excellence; others commented on the demanding

expectations they had encountered, for example, "producing high levels of excellence in attitudes and achievements." There was also concern about how to deal with the drive to achieve high results on standardized tests, particularly in apposition to developing students' individual potentials and meeting the rapidly changing demands of society on both economic and technological fronts.

Governance issues. Ranked *most important* by 8% of the respondents, these issues related to political action, particularly in connection with minority needs and public versus private schooling.

For example, some directors in Ontario had been drawn into the political and legal skirmishes that had been precipitated by the introduction of government funding of Roman Catholic schools for Grades 11 and 12, while one principal was engaged in efforts to ensure that such funding would not dilute the *catholicity* of separate schools, "the only justification for a segregated school." Another director reported the existence of several Christian schools in his jurisdiction and was trying "to work out a compromise that would keep the schools together."

Other issues in this category included the need to develop satisfactory governance structures for the schooling of Native students, the challenge of developing a strong board, and the introduction of heritage languages in a time of fiscal restraint. One of the British Columbia principals was concerned with changing the attitudes of a provincial government that placed a low emphasis on education.

Managerial issues. Ranked *most important* by 8% of the respondents, these issues related to decision making, mediation, or general management aspects of administration. Specific areas of concern included difficult administrator-trustee relationships; ensuring that programs were being properly implemented; quality of communication between the administrator and his or her students, staff, and parents; and responsibilities so broad that effectiveness was compromised.

Student personnel issues. Issues of this type, ranked *most important* by 7% of the respondents, related directly to student behavior. The responses focused on attendance, discipline, enrichment for capable but underachieving students, ensuring that schools were preparing students for the future, and on trying to save students from "a life of chronic unemployment or short-term units of employment."

Reasons for Assigning High Priority

In general the imperatives that these principals and superintendents reported had three roots: a service ideology, sociopolitical forces, and personal belief systems.

Service ideology. The largest class of reasons for assigning highest priority reflected a strong commitment to improving and protecting the quality of education provided for students:

> *It's obvious: our function is to educate children and young men and women and to do it in the best possible manner; [all of] what we do facilitates this priority.* (director)

> *The shortage of financial resources ... could mean cutting programs.... [However] I've always been an advocate of doing the most we can for the most kids.... I want to avoid cutting back on or eliminating programs or services.* (principal)

Sociopolitical forces. The next most common set of reasons identified pressures from the external environment such as student demographics; the priorities of political parties; demands expressed by trustees, parents, and business people; the imperatives of the job market and technological developments; and the need to compete with other agencies for limited resources including students. For example, one superintendent observed:

> *Preparing students for a tenuous future is very difficult. Pressure groups and self-interest oriented individuals are trying to have it go their way. Many people expect the schools to do more and to do it better. At the same time the political pressure in the province appears to be moving in the direction of more funding for private schools and [separate] schools.*

Personal beliefs. The third clear class of reasons, albeit small in number, was rooted in strong personal beliefs or values. For example:

> *I am strongly against the pervasive "regression to the mean" mentality which threatens to engulf our educational system. What's hot is mediocrity; what's not is meritocracy.* (principal)

> *I believe that we need ... [to move] from the confrontational mode to a mutual problem solving mode.... [There is] enormous waste in the confrontational mode: the time, energy, and stress on both sides of the confrontation and ... the positive and productive ideas that never come to the surface from either side.* (director)

When the reasons for assigning highest priority were cross-tabulated by office/role to ascertain whether principals and superintendents differed in their orientations to the environment, the expected intra/extra-organizational differential did not emerge.

Discussion

We started by noting that strong support is found in the literature for what we have termed the congruity imperative—the need to maintain *fit* between organizational structures and environmental conditions—and the importance of administrators systematically monitoring the environment for changes and effecting appropriate adjustments in how the organization functions. We also noted that a review of the literature provided little information on the extent to which school administrators pay attention to the congruity imperative. We therefore turn now to a consideration of the ways in which the research reported here has added to information on the congruity imperative in school administration. But before we proceed, a note of caution is in order. Given the select group that participated in this survey and the exploratory nature of this study, the conclusions reached must be considered tentative.

Are Principals and Superintendents Most Concerned About "Maintenance" or "Adaptive" Matters?

As noted earlier, our data indicate that the principals and superintendents in the sample most frequently gave the highest priority to issues associated with teaching personnel, finances, program, governance, management, and

student personnel. More important, the details of their responses consistently indicated that they attached greatest importance to activities that would have the effect of maintaining or improving the quality of education and students' welfare. We conclude that in the data on highest priority issues, the dominant orientation was intra-organizational and of the *maintenance* type (Tosi & Carroll, 1976). Thus our data seem to support the following investigators' conclusions: Larson et al. (1981)—Superintendents tend to focus on internal management; Friesen and Duignan (1980)—A very large proportion of a superintendent's *contact* time is spent with *internal* persons/groups, very little with *externals*; Gunn (1984)—Principals prefer to work at improving conditions for students and teachers, improving the school's effectiveness by developing harmony and high staff morale; McLeod (1984)—Directors and superintendents tend to focus on orderly internal functioning and "the prevention of nasty disruptions"; they tend to adopt a *functionary* role.

To What Extent Are Principals' and Superintendents' Priorities Influenced by Factors in the External Environment?

The literature indicates that effective executives actively scan the environment to identify forces and trends that may be of consequence to their organizations. This study provided little evidence that these reputedly effective administrators scanned the environment with an anticipatory set; their high priority activities seemed to focus more on present demands than on understanding and preparing for the future. Furthermore, the imperatives that underscored the respondents' highest priorities were dominated by the professional ideology (Evan, 1973; Fris, 1975) of providing the best service possible. Like the third cluster or category of imperatives (personal beliefs), this group of reasons cannot be characterized easily as either extra- or intra-organizational, if only because ideologies (especially the professionalism ideology) are often both personalized and shared widely within and without an organization. We seem to have here confirmation of McPherson, Crowson, and Pitner's (1986) caution that "the institutionalized rules, norms, and myths of the larger society are fully reflected in the ... structure of an organization" (p. 177) and that the notion of an organization-environment demarcation has its limits.

The second largest cluster of imperatives for assigning highest priority to an issue—sociopolitical forces—clearly reflected an awareness of and a desire to respond to extra-organizational factors. These data lend support to Payne's (1987) finding that principals are sensitive to parents' concerns and use parents as monitors of the school's various environments. However, our respondents did not evidence the selective responsiveness that was apparent in the studies by Grainger (1984) and Payne (1987). The data reflected a willingness to respond to a variety of pressures from a wide variety of individuals and groups in the external environment.

Finally, as noted before, the expectation that principals would focus primarily on in-house issues while superintendents would focus on issues that originate outside the school system was not supported; superintendents

and principals appeared to be responding to issues in both the internal and external environments of their organizations.

The Congruity Imperative

The data from the exploratory study reported here provide support for the notion that the activities of administrators are indeed influenced by forces in the so-called external environment. However, our data also suggest that the concept of a congruity imperative can be oversimplified.

In its most elementary form, the congruity principle posits that activities, procedures, and structures in an organization must be made to *fit* the exigencies of the environment if that organization is to be successful or even to survive. However, in the literature there are three competing models of the organization/environment relationship (McPherson et al., 1986, pp. 180-200).

Organizational ecology perspective. In this perspective, "it is the environment which is ... the key determinant of organizational form.... It is the terms of the environment that must be honored first and foremost" (pp. 199, 203). In this perspective "the shrewd administrator provides a regular opportunity for ... the inclusion of all points of view in matters that affect children and the school" (p. 205), and internal control and effectiveness are seen to come from responsiveness to the environment.

Public choice or political economy perspective. In this perspective organizations are seen to be shaped by the self-interested actions of individuals; the individuals who constitute an organization maximize their own welfare by pursuing only those rewards (both within and external to the organization) that they idiosyncratically value. An administrator who holds this perspective will concentrate first on understanding the rewards that motivate the organization's members, then on the managerial implications.

Political systems perspective. In this view of organizations there is "a constant dialectic of adaptation between organization and environment—a continuing struggle toward equilibrium [through] a continuous give and take, a mutuality of input and output, of feedback and adjustment" (pp. 186, 201).

Although we do not report our respondents' descriptions of how they were dealing with their most important challenges, we can say that those data support the generalization that "organizations do not engage their environments successfully through a simple absorption of forces from their surroundings.... Interaction is a two-way boulevard.... Both sides (organization and environment) reach out, both sides change, both sides adapt" (McPherson et al., 1986, pp. 182-183). Moreover, those data indicate that all three of the above models must be employed to adequately account for the qualitative variability in organization/environment interactions.

One of the important conclusions is that the three models of organization/environment link may be viewed as descriptive of three archetypical styles or approaches to the management of organizational adaptation. And one might hypothesize that school administrators are more effective if they

are skilled in the approach(es) that is (are) appropriate to the exigencies of the situation.

Future Research

The above speculations lead to two important implications for research on administrative effectiveness. The first is that investigators should continue to look for distinctive patterns of administrative behavior that are associated with organizational adaptation—there is a need to test the robustness of the threefold typology of responses to the congruity imperative. The second research challenge is to determine the conditions (contingent circumstances) under which each of various approaches to the management of the organization/environment synapse is most appropriately invoked.

Note

1. Assistance with the data analysis was provided by four Master's students in the Department of Educational Administration at the University of Alberta: Pat Campbell, Florence Glanfield, Marion Holder, and Karran Thorpe.

References

Aguilar, F.J. (1967). *Scanning the business environment.* New York: Macmillan.

Burns, T., & Stalker, G. (1961). *The management of innovation.* London, UK: Tavistock.

Crowson, R.L. (1987). The local school district superintendency: A puzzling administrative role. *Educational Administration Quarterly, 23*(3), 46-69.

Cuban, L. (1985). Conflict and leadership in the superintendency. *Phi Delta Kappan, 67*(1), 28-30.

Evan, W.M. (1973). Some occupational implications for designing an experimental program in educational administration. *Journal of Educational Administration, 11,* 216-243.

Feilders, J.F. (1978). *Action and reaction: The job of an urban superintendent.* Unpublished doctoral dissertation, Stanford University.

Friesen, D., & Duignan, P. (1980). How superintendents spend their working time. *Canadian Administrator, 19*(5), 1-5.

Fris, J. (1975). The professional status of teachers: Some theoretical concerns and an Ontario survey. *Canadian Administrator, 14*(5), 1-5.

Grainger, R.A. (1984). *The participation of parents in local school decision making.* Unpublished doctoral dissertation, University of Alberta.

Gunn, J.A. (1984). *Job satisfaction of senior high school principals and their perceptions of school effectiveness, their leadership, and their bases of influence.* Unpublished doctoral dissertation, University of Alberta.

Hannaway, J., & Sproull, L.S. (1978-79). Who's running the show? Coordination and control in educational organizations. *Administrator's Notebook, 30*(6), 1-4.

Katz, D., & Kahn, R. (1966). *The social psychology of organizations.* New York: Wiley.

Larson, L.L., Bussom, R.S., & Vicars, W.M. (1981). *The nature of a school superintendent's work: Final technical report.* Carbondale, IL: Southern Illinois University.

Lawrence, P.R., & Lorsch, J.W. (1967). *Organization and environment: Managing differentiation and integration.* Harvard, MA: Graduate School of Business Administration, Harvard University.

McLeod, G.T. (1984). The work of school board chief executive officers. *Canadian Journal of Education, 9*(2), 171-190.

McPherson, R.B., Crowson, R.L., & Pitner, N.J. (1986). *Managing uncertainty: Administrative theory and practice in education.* Columbus, OH: Merrill.

Mintzberg, H. (1973). *The nature of managerial work.* New York: Harper & Row.

Morris, V.C. (1984). *Principals in action: The reality of managing schools.* Columbus, OH: Merrill.

Morrison, J.L., Renfro, W.L., & Boucher, W.I. (Eds.). (1983). *Applying methods and techniques of futures research.* New Directions for Institutional Research Number 39. San Francisco: Jossey-Bass.

Payne, C.H. (1987). *Parents and control of elementary school principal work behavior*. Unpublished doctoral dissertation, University of Alberta.

Peters, T.J., & Waterman, R.H. (1982). *In search of excellence: Lessons from America's best-run companies*. New York: Warner.

Peterson, K.D. (1984). Mechanisms of administrative control over managers in educational organizations. *Administrative Science Quarterly, 29*(4), 573-597.

Pitner, N.J. (1978). *Descriptive study of the everyday activities of suburban school superintendents: The management of information*. Unpublished doctoral dissertation, Ohio State University.

Pitner, N.J., & Ogawa, R.T. (1981). Organizational leadership: The case of the superintendent. *Educational Administration Quarterly, 17*, 45-65.

Renfro, W.L., & Morrison, J.L. (1983). Anticipating and managing change in educational organizations. *Educational Leadership, 41*(1), 50-54.

Thompson, J. (1967). *Organizations in action*. New York: McGraw-Hill.

Tosi, H.L., & Carroll, S.J. (1976). *Management: Contingencies, structure, and process*. Chicago: St. Clair.

Woodward, J. (1965). *Industrial organization*. London: Oxford University Press.

The Alberta Journal of Educational Research *Vol. XXXIV, No. 4, December 1988, 390-402*

CHERYLL DUQUETTE

and

ROBERT R. O'REILLY
University of Ottawa

Perceived Attributes of Mainstreaming, Principal Change Strategy, and Teacher Attitudes Toward Mainstreaming

The attitudes toward mainstreaming held by elementary teachers who had some experience integrating exceptional students in their regular classes were examined. These attitudes were related to the perceived attributes of mainstreaming. Rogers' revised model of innovations was used to identify the attributes. Attitudes were also related to the behavior of the school principal with respect to mainstreaming using the model of Dalin. The sample consisted of 189 experienced teachers from three large boards of education in an urban area in eastern Ontario. Results show that teachers' attitudes are generally but not universally favorable toward mainstreaming. Attributes of mainstreaming perceived by the teachers support Rogers' model. Dalin's model related to principal behaviors is also supported.

One of the most widely implemented innovations in education during the past decade has been the move to integrate disabled children into the educational services. This means at one end of the process the closing down of segregated educational facilities, and at the other the placement of exceptional children in the regular classroom. These innovations are part of a general movement toward *deinstitutionalization* of the disabled and the *normalization* of their treatment by service providers.

Legislative changes in Canada (O'Reilly, 1984) have made the local public school system the major agent responsible for the provision of educational services for the disabled. The resulting policies and practices have strongly encouraged the practice of mainstreaming, which is generally defined as the placement of an exceptional child in a regular classroom under the direction of the regular teacher for a significant portion of the child's educational program.

The attitudes of teachers toward any innovation determine in part their motivation to attempt the change, or once having made the attempt, to per-

---

Dr. Duquette is a professor of teacher education with extensive experience in the field of special education.

Dr. O'Reilly is a professor of educational administration with interests in the management and evaluation of services for exceptional children.

severe in or discontinue the practice. A variety of recent studies has shown that positive attitudes of teachers contributed to the success of mainstreaming (Brooks & Bransford, 1971; Larrivee, 1981; Roubinek, 1978). Yet not all teachers have positive attitudes. Although some studies have shown that teachers and teachers in training do have positive attitudes toward mainstreaming (Curtis, 1985; Harasymiw & Horne, 1975; Higgs, 1975; Winzer, 1984, 1987), other investigators have reported that teachers have some reservations about it (Barngrover, 1971; Bond & Dietrich, 1982; Bradfield, Brown, Kaplan, Rickert, & Stannard, 1973; Shotel, Iano, & McGettigan, 1972). Moreover, Crealock and Nediger (1983) reported that elementary teachers in Ontario held neutral attitudes toward mainstreaming before they actually experienced it.

This article describes a study which investigates the correlates of attitudes toward mainstreaming of teachers who have already implemented the innovation. For this study, mainstreaming refers to the instruction of exceptional students in regular education classes for at least 25% of the day. Children identified as gifted are not included in the definition of exceptional in this study. In particular, it addresses two questions with respect to elementary level teachers who have had some experience integrating exceptional children in their regular classrooms:

1. What are the perceived attributes of mainstreaming which affect teachers' attitudes toward mainstreaming?
2. Which behaviors of school principals affect teachers' attitudes?

Theoretical Framework

In many school districts in the United States and Canada, mainstreaming is considered innovative (Keeton, 1982; McLaughlin-Williams, 1983; Roubinek, 1978; Spodek, 1982; Stannard, 1976). Teacher concerns about educational innovations are instrumental to the success or failure of implementation. These personal concerns emphasize the notion that in any change of effort "the human element is of prime importance in the acceptance or rejection of an innovation" (Blank & White, 1984, p. 73). Teacher attitudes toward the innovation of mainstreaming is considered crucial (Winzer, 1987), and it appears that other variables such as teacher age, grade level taught, and the nature of the child's disability may be related to attitude formation (Antonak, 1980; Berryman & Berryman, 1981; Larrivee, 1982). Two more variables that may also be related to teacher attitudes are perceptions of the characteristics of mainstreaming and the principal's behavior during the implementation of mainstreaming. The study reported in this article focuses on teacher attitudes toward mainstreaming as related to teacher perception of the characteristics of mainstreaming and to the behavior of the school principal.

Teacher Perception of Mainstreaming

The role of the perceptual process in attitude formation toward innovation is emphasized by Sherif and Cantril (1945). Perception, they report, is al-

ways involved because attitudes are derived from experience rather than being innate.

A recent survey of the research on perception (Johnson, 1987) provides some recommendations and cautions for educational researchers. Although perceptions are dominant and critical determinants of behavior, they are subject to a variety of cultural, personal, organizational, and other factors which attenuate an individual's ability to perceive accurately. However, Johnson also suggests that there is considerable consistency in social perceptions among members of a group. Rogers and Shoemaker (1971) related perceptions to attitude formation and the decision to accept or reject an innovation in their model of the innovation decision process. They identify five perceived characteristics of an innovation which are salient in this process: the relative advantage of the innovation; the compatibility of the innovation with current practices and values; the complexity of the innovation and the difficulty of implementing it; the possibility of attempting a trial run without committing oneself totally to the innovation; and the observability and visibility of the innovation. Their model provided the basis for a study conducted by McLaughlin-Williams (1983). She reports that the best predictors of teacher attitudes are beliefs concerning relative advantage, adaptability, and observability of the innovation of mainstreaming. However, she notes that dimensions describing each of the attributes are not discrete and suggests caution in the interpretation of results as well as a need for further studies in the area of teacher perceptions of mainstreaming.

Recently, Rogers (1983) revised his model of the attributes of an innovation which influence a person's behavior in the innovation decision process. In the updating, the five perceived characteristics of an innovation are more precisely defined by the incorporation of 11 dimensions which provides a clearer description of the characteristics than the Rogers and Shoemaker (1971) version.

In the revised model, relative advantage of an innovation is composed of four dimensions: economy of effort and use of resources; social prestige accruing to the implementor; convenience; and satisfaction. Compatibility of an innovation refers to the degree to which the innovation is perceived as being consistent with the existing value system, with the teacher's previous experience, and with the needs of the teacher and students. Complexity has two dimensions: the innovation viewed as easy or difficult to perform, and easy or difficult to understand. The remaining characteristics are each composed of one dimension. Trialability is the extent to which any innovation may be tried gradually; observability is the extent to which an innovation is visible to others (Rogers, 1983, pp. 213-232).

Administrative Support for Mainstreaming

Administrative support for educational innovations has also been linked to successful implementation (Barnes, 1970; Berman, Greenwood, McLaughlin, & Pinens, 1975; Daft & Bekker, 1978; Dalin, 1978). Indeed, Nicholsen and Tracy (1982) suggest that in the implementation of an edu-

cational innovation, the principal's actions contribute to teacher attitudes toward the innovation. Not surprisingly, administrative support has also been shown to be related to the success of the mainstreaming (Sivage, 1980). Moreover, administrative support has been demonstrated to be related to teacher attitudes toward mainstreaming (Guerin & Szatlocky, 1974; Larrivee & Cook, 1979). Sivage (1980) states that certain school behaviors of the principal are related to the successful implementation of mainstreaming. However, it appears that some exogenous variables may also be included in the principal's role behaviors related to teacher attitudes toward mainstreaming. It seems that the principal plays a liaison role between the school and the parents of exceptional students (Rocha, 1983). In addition, the principal should at least be aware of the extent to which teacher unions support the implementation of mainstreaming (Zey, 1981). At present, there exists a paucity of studies in which a theoretical framework incorporating the multitude of endogenous and exogenous variables related to the principal's role behavior during the implementation of mainstreaming has been utilized.

Dalin's (1978) model of educational change related to the role of the educational administrator in implementing change was used as a framework for collecting teachers' perceptions of the principal's role in implementing mainstreaming in the schools. Using Dalin's model, the administrator's behavior may be divided into four factors: the educational setting, the environment, the innovation, and the change strategy. In a first adaptation of the model for educational research, Duquette (1983) found that each of the four factors is related to the implementation of curricular innovations in Ontario schools. These factors have the added advantage of focusing more attention on the boundary-spanning roles of the principal and the exogenous variables in the change process than do most other models of educational change which tend to focus more on processes internal to the school.

It was therefore the purpose of this study to examine the relationship of teacher perceptions of mainstreaming and the principal's role behaviors with attitudes of regular education teachers toward mainstreaming. Specifically, the 11 dimensions of Rogers' (1983) perceived characteristics of an innovation and Dalin's (1978) four factors related to educational change were assessed on predictions of attitudes toward mainstreaming. The general hypothesis tested in this study was that teacher perceptions of mainstreaming and teacher observation of the principal's behavior are related to teacher attitudes toward mainstreaming. Two specific hypotheses were also tested. It was predicted that each of Rogers' (1983) 11 dimensions of perceived characteristics of innovation as applied to mainstreaming is related to teacher attitudes toward mainstreaming. It was further hypothesized that each of Dalin's (1978) four factors of educational change as applied to the principal's role behaviors during the implementation of mainstreaming is related to teacher attitudes toward mainstreaming.

Research Methodology

Sample

The sample consisted of English-speaking teachers from selected K-6 schools in three large boards of education in eastern Ontario. Using alphabetized lists of schools provided by the three boards of education, every second K-6 school was selected. This procedure resulted in a total sample of 47 schools employing approximately 600 teachers, only a fraction of whom were potential respondents. The eligible sample was limited to only those teachers who had integrated exceptional students (not gifted) for at least 25% of the school day.

Over one half of the teachers in the sample were between 36 and 45 years of age, and 31% were teaching grade four or five. Fifty-six percent of the teachers mainstreamed exceptional students who remained in special education classes for most of the school day, while 34% worked with students who leave the regular classrooms for short periods to receive assistance from a resource teacher. Only 10% of the respondents taught exceptional students for the entire school day in the regular classroom.

The average number of exceptional students mainstreamed into one classroom was 1.3. Teachers in this sample appeared to be integrating a broad range of exceptional children: 23% of the children were mentally handicapped, 35% had behavioral problems, 20% had communication disorders, and 12% had severe physical handicaps. The remaining 10% were children with multiple handicaps.

Although over half the teachers in this sample reported that they had mainstreamed for two or more years, only 40% reported completing any courses in special education.

Instrumentation

The instrument for this study was a four-part teacher questionnaire consisting of 95 items. The first section consisted of 12 items designed to elicit demographic information and to screen respondents.

The second part of the questionnaire consisted of 33 items measuring teacher perceptions of mainstreaming. The instrument was based on a survey used by McLaughlin-Williams (1983) which was revised to reflect Rogers' (1983) updated model. Items in this section were field-tested to establish content and construct validity.

Items in the third part of the survey were adapted from an existing questionnaire (Duquette, 1983) which measured perceptions of the principal's role behaviors when implementing a new curriculum document. Dalin's (1978) factors of educational change were also utilized as the theoretical framework for this survey. As new curricula and mainstreaming were both considered educational innovations and because Dalin's (1978) conceptualization was used in both studies, it seemed that the previous instrument would provide an appropriate basis for this research. The wording in many of the items was changed slightly to reflect mainstreaming rather than curriculum implementation. The 37 items in this section of the questionnaire were also field-tested to establish content and construct validity.

The final section of the teacher questionnaire contained eleven semantic differential-type items, a demographic item, and an item related to teacher attitudes toward mainstreaming. Although there were three measures of teacher attitude, items in the semantic differential served as the dependent variable for this study.

Field-testing of the four-part questionnaire was conducted with graduate students enrolled in a research design course at the University of Ottawa. These graduate students were also elementary school teachers who had mainstreamed exceptional children in their regular education classrooms. No changes were made to the questionnaire after field-testing.

Data Collection

Packets containing instructions, questionnaires, and return envelopes were sent to each of the 47 selected schools. Principals were telephoned to ensure that the packets had arrived at the school and to clarify any instructions. Although most surveys were returned within 14 days, follow-up telephone calls were made four weeks later to the few schools from which teacher questionnaires had not been received.

Data Analysis

Data were analyzed using Pearson product moment correlation coefficients to examine the relationship of each of the 11 dimensions of characteristics of an innovation and the four factors of educational change to teacher attitudes toward mainstreaming. To determine which of the dimensions and factors were the best predictors of teacher attitudes, stepwise regression analyses were done using the 11 dimensions as predictors of attitudes toward mainstreaming in one analysis and the four factors of educational change in the other.

Results

The final sample consisted of 189 teachers who had experienced mainstreaming of exceptional students for at least 25% of the day. Although the total number of teachers who mainstream in the selected schools is unknown, it was assumed that the 189 represented a substantial proportion. However, it is acknowledged that the threat of a biased sample may have existed.

Teacher Perceptions of Mainstreaming

Teacher perceptions of the various attributes of mainstreaming were measured using a 33-item scale which formed the second part of the questionnaire. Responses to items were scored in a Likert-type fashion from 0 to 4.

Internal reliability for each of the 11 dimensions of the scale, as calculated with Cronbach alpha coefficients, ranged from .18 to .69. While some of these coefficients are quite small, the results are not surprising because each dimension was measured by only three items. Pearson product moment correlation coefficients were calculated between each of the dimen-

sions of perception and attitude toward mainstreaming (Table 1). Coefficients range from .25 to .53 and all but two were greater than .35.[1]

Each of the 11 dimensions correlated significantly with attitudes toward mainstreaming, which provides support for Rogers' (1983) revised model. A stepwise regression analysis was completed to determine which set of dimensions was the best predictor of attitude. The BMDP2R stepwise regression package was used to conduct this analysis. As shown in Table 2, the set of dimensions which is the best predictor of attitude includes *difficult to do, observability, needs, existing values, trialability, social prestige,* and *economy*. These dimensions are representative of all five of Rogers' attributes of an innovation.

Teacher Observations of the Principal's Behaviors

The principal's behaviors as perceived by the teachers were measured using a 37-item instrument which formed the third part of the teacher questionnaire. Responses to items were in a no/yes format and were scored 0 and 1 respectively.

Internal reliability was calculated with the Kuder-Richardson (formula 20) test. KR coefficients range from .62 to .84. Pearson product moment correlation coefficients between each of the factors of principal behavior

Table 1
Correlation Between Teacher Attributes of Mainstreaming and Attitudes Toward Mainstreaming

Teacher Attributes of Mainstreaming	r
Relative Advantage	
Economy	.38
Social prestige	.30
Convenience	.41
Satisfaction	.40
Compatibility	
Existing values	.39
Past experience	.25
Needs	.52
Complexity	
Difficult to understand	.36
Difficult to do	.53
Trialability	
Trialability	.39
Observability	
Observability	.42

N = 189. All rs are significant ($p < .01$).

Table 2

Stepwise Regression Using 11 Dimensions of Teacher Attributes of Mainstreaming as Predictors of Attitudes Toward Mainstreaming

Dimension	Multiple R	R Squared	Regression Coefficient
Difficult to do	.53	.28	1.08
Observability	.61	.37	1.33
Needs	.66	.43	.72
Existing Values	.68	.46	.99
Trialability	.69	.48	.72
Social Prestige	.70	.49	.77
Economy	.71	.51	.86

$F(7, 182) = 26.61$, $p < .01$

and teacher attitudes toward mainstreaming ranged from .28 to .37. All correlations are statistically significant (Table 3), which provides support for Dalin's (1978) model. Because all factors are statistically related to attitudes, a stepwise regression analysis was performed to determine which set of factors was the best predictor of attitudes. The BMDP2R stepwise regression package was again used. The factor of innovation is the only predictor to emerge from this analysis. This suggests that there is considerable common variance among the four factors.

Perceptions of Mainstreaming and the Principal's Behavior

Using the BMDP2R stepwise regression package, an analysis was done to determine the set of dimensions and factors that was the best set of predictors of attitudes toward mainstreaming. For this set of data, the best predictors of teacher attitudes are the perceptions *difficult to do, observability, existing values, convenience*, the principal behavior factor of *innovation*, and the perception *satisfaction* (Table 4).

Teacher Attitudes

Items in the semantic differential questionnaire measured teacher attitudes toward mainstreaming. The Cronbach *alpha* for the scale was .90. Respon-

Table 3

Correlations Between Teacher Perceptions of Principal Behaviors and Attitudes Toward Mainstreaming

Teacher Perceptions of Principal Behaviors	r
Educational setting	.28
Environment	.30
Innovation	.37
Change strategy	.33

$N = 189$. All rs are significant ($p < .01$).

Table 4
Stepwise Regression Using 11 Dimensions of Teacher Attributes of Mainstreaming and Four Factors Measuring Principal Behaviors as Predictors of Attitudes Toward Mainstreaming

Dimension	Multiple R	R Squared	Regression Coefficient
Difficult to do	.57	.33	.58
Observability	.64	.41	.29
Existing Values	.68	.45	.22
Convenience	.69	.49	.20
Innovation	.72	.52	.17
Satisfaction	.73	.53	.16

$F(6, 166) = 31.69$, $p < .01$

ses to the items were scored from 0 to 6, and for most items the scores tended to cluster on the positive half of the scale, indicating that the attitudes toward mainstreaming of this group of experienced teachers tend to be positive. Standard deviations for each item tend to be large, which suggests that the instrument did discriminate between various teacher attitudes.

Although the semantic differential was used to test the hypothesis of the study, two other items in the data collection package also provided information about teacher attitudes toward mainstreaming. Responses to a global item measuring teacher attitude toward the experience of mainstreaming were scored on a scale from 0 to 2. The mean 1.34 suggests that the teachers felt positive about their own experiences with mainstreaming. The second item asked teachers if they would freely choose to mainstream the following year. On a two-point scale of no/yes (scored 0 or 1), this sample's mean was 0.76, indicating that more than three quarters of the teachers would voluntarily choose to mainstream children in their classes again the following year. Thus the analysis of these two items, as well as the analysis of the semantic differential, suggests that many teachers feel positive about mainstreaming.

Discussion and Summary

The results of this study demonstrate that a sample of all teachers in the major school boards of the Ottawa area who have mainstreamed exceptional children in their classes generally have positive attitudes toward mainstreaming. These attitudes are correlated with the perceptions that teachers have about the nature of mainstreaming (its costs, benefits, ease of implementation, burdens, etc.). As well, the attitudes are correlated with the teachers' perceptions of the behavior of the school principal in support of mainstreaming both inside and outside the school.

Experienced teachers who have positive attitudes, as measured by the semantic differential, perceive that the day-to-day activities required by mainstreaming such as adapting classroom programs and completing extra

forms are not burdensome. They do not perceive that mainstreaming requires a lot of extra time and work on their part.

Positive teacher attitudes toward mainstreaming are also related to a perception that the regular classroom placement of an exceptional child contributes to positive changes in the pupil's academic performance and/or behavior. These improvements should be visible to the mainstreaming teacher, other teachers, and parents. A third element of perception which is related to positive attitudes toward mainstreaming is its compatibility with current values and professional beliefs held by educators regarding the placement of exceptional children. Essentially, these teachers perceive that the requirements of mainstreaming are easily managed, that integration contributes to improved pupil performance, that the philosophy of mainstreaming is compatible with their own beliefs, and that they generally feel satisfied with the implementation of mainstreaming.

The behavior of school principals also seems to be related to positive attitudes toward mainstreaming. Associated with positive teacher attitudes are the following behaviors of principals. The principal makes clear the rationale and objectives of mainstreaming; clarifies the roles of teachers, consultants, and themselves; and makes known the support of mainstreaming by senior administrative officials. In addition, the principal maintains a flexible organization to meet the requirements of mainstreaming and gives attention to solving problems which individual teachers may have concerning mainstreaming. In summary, teacher observations of the active support and commitment to mainstreaming by the school principal appear to be related to a teacher's attitude toward mainstreaming. Data from this study suggest that exogenous factors, that is, behaviors of the principal which occur or are related to agencies outside the school, are not as strongly related to teacher attitudes as are actions of the principal which concern activities inside the school. Principal behaviors which require face-to-face interactions in the school appear to be more important in this regard than behaviors associated with the principal's role at the board office or in the community.

These sets of results, therefore, provide support for the theoretical framework of this study. For this sample, all 11 dimensions of Rogers' (1983) perceptions of an innovation are positively related to attitude as measured by the semantic differential scale and the *willingness to mainstream* item. The data not only provide support for the revised model of the innovation decision process but also attest to the utility of the measuring instruments. As well, all four factors related to the principal's behaviors produced significant positive correlations with attitudes toward the innovation of mainstreaming. These findings, therefore, provide support for the second element of the theoretical rationale, Dalin's (1978) conceptualization of educational change. However, because there appears to be a lack of discrimination among the four factors, researchers should be cautioned before conducting further research with Dalin's model. Certainly, scales with greater power to discriminate between factors are required. It is also possible that the descriptions of the factors by Dalin need to be refined.

What can be learned from the perceptions of these experienced teachers? While the majority of teachers seemed to have a positive attitude toward mainstreaming, it must be recognized that not all teachers found mainstreaming exceptional children to be a rewarding experience. Almost one quarter of them said that if possible they would avoid mainstreaming in the future. Some teachers did not view the attributes of mainstreaming in a positive light. For some, the rewards were few and the effort great.

Educational leaders who wish that all teachers support mainstreaming can look to the attributes of mainstreaming suggested by this study to be directly related to positive attitudes.

Johnson (1987) provides some direction to educational leaders who wish that all teachers support mainstreaming. Organizational factors such as leader behavior and expectations, systems for rewards, and allocation of resources are important organizational factors which influence teacher perceptions. Administrators can ensure that school policies, regulations, and resources support teachers who have exceptional children in their classroom. They can reduce the perceived difficulty of the task by providing suitable inservice training for teachers (Hummel, Dworet, & Walsh, 1985). Finally, they can engage in those innovation implementation behaviors suggested by Dalin.

In conclusion, mainstreaming is a reality in many school districts in the United States and Canada. It is generally accepted that positive teacher attitudes contribute to the success of mainstreaming of exceptional students in the school. The results of this study indicate that teacher perceptions of mainstreaming and the behavior of school principals are related to teacher attitudes toward mainstreaming. Therefore, consultants involved in the implementation of mainstreaming should attempt to develop positive perceptions of the characteristics and advantages of mainstreaming among regular education teachers. In addition, administrators should recognize their role in providing support to teachers implementing mainstreaming so that positive attitudes toward integration, a key element in the effectiveness of mainstreaming, are developed.

Acknowledgments

Suggestions and comments by W.H. Worth and two anonymous reviewers assisted the authors in the preparation of this article.

Note

1. Detailed information concerning the questionnaires used in this study and the results obtained are available from Dr. Cheryll Duquette, Faculty of Education, University of Ottawa, 651 Cumberland, Ottawa, ON, K1N 6N5.

References

Antonak, R. (1980). A hierarchy of attitudes toward exceptionality. *Journal of Special Education, 14*, 231-241.

Barnes, M. (1970). The administrator's role in humanizing the school. *National Elementary Principal, 40*(4), 37-40.

Barngrover, E.A. (1971). A study of educators' preferences in special education programs. *Exceptional Children, 37*, 754-755.

Berman, P., Greenwood, P., McLaughlin, M., & Pinens, J. (1975). *Federal programs supporting educational change* (Vol. 5). Santa Monica, CA: Rand.

Berryman, J., & Berryman, C. (1981). *Use of the Attitudes Toward Mainstreaming Scale with rural Georgia teachers*. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles.

Blank, S., & White, R.B. (1984). Teachers and technological innovation. *B.C. Journal of Special Education, 8*, 71-80.

Bond, C.L., & Dietrich, A. (1982). *Regional survey of teacher attitudes toward special education resource programs*. (ERIC Document Reproduction Service, No. ED 228 772).

Bradfield, H., Brown, J., Kaplan, P., Rickert, E., & Stannard, R. (1973). The special child in the regular classroom. *Exceptional Children, 39*, 384-390.

Brooks, B., & Bransford, L. (1971). Modification of teachers' attitudes toward exceptional children. *Exceptional Children, 38*, 259-260.

Crealock, C., & Nediger, W. (1983). *Bill 82 and you*. Toronto: Ontario Public School Teachers' Federation.

Curtis, C. (1985). Education students' attitudes toward disabled persons and mainstreaming. *Alberta Journal of Educational Research, 31*, 288-305.

Daft, L., & Bekker, B. (1978). *Innovations in organizations*. New York: Elsevier North-Holland.

Dalin, P. (1978). *Limits to educational change*. Oslo: Macmillan.

Duquette, C. (1983). *The principal's role in curriculum implementation questionnaire: The development and validation of an instrument to measure Dalin's Model of Educational Change*. Ottawa: Author.

Guerin, G., & Szatlocky, K. (1974). Integration programs for the mildly retarded. *Exceptional Children, 41*, 637-643.

Harasymiw, S., & Horne, M. (1975). Integration of handicapped children: Its effect on teacher attitudes. *Education, 96*, 153-158.

Higgs, R. (1975). Attitude formation—contact or information? *Exceptional Children, 41*, 496-497.

Hummel, J.W., Dworet, D., & Walsh, M. (1985). Exceptional students in regular classrooms: Teacher attitudes and teacher inservice needs. *Canadian Journal for Exceptional Children, 2*, 14-17.

Johnson, N. (1987). The pervasive, persuasive power of perceptions. *Alberta Journal of Educational Research, 33*, 206-228.

Keeton, A. (1982). Teacher liability. *Ontario Education, 14*, 12-15.

Larrivee, B. (1981). Effect of inservice training intensity on teachers' attitudes toward mainstreaming. *Exceptional Children, 48*, 34-39.

Larrivee, B. (1982). Factors underlying regular classroom teachers' attitude toward mainstreaming. *Psychology in the Schools, 19*, 374-379.

Larrivee, B., & Cook, L. (1979). Mainstreaming: A study of the variables affecting teaching attitude. *Journal of Special Education, 13*, 315-324.

McLaughlin-Williams, J. (1983). *Innovation, diffusion, adoption: Regular teachers' perceptions of mainstreaming*. Unpublished doctoral dissertation, University of Michigan.

Nicholsen, E.W., & Tracy, S.J. (1982). Principal's influence on teacher's attitude and implementation of curricular change. *Education, 103*, 68-73.

O'Reilly, R.R. (1984). Educational rights for disabled children. *Just Cause, 2*(3), 4-6.

Rocha, R. (1983). Educating the handicapped—The need for close cooperation between parents, administrators. *NASSP Bulletin, 67*, 92-97.

Rogers, E.M. (1983). *Diffusion of innovation* (3rd ed.). New York: Free Press.

Rogers, E.M., & Shoemaker, F. (1971). *Communication of innovations: A cross cultural approach*. New York: Free Press.

Roubinek, D. (1978). Viewpoint: Will mainstreaming fit? *Educational Leadership, 35*, 410-412.

Sherif, M., & Cantril, H. (1945). The psychology of "attitudes." *Psychological Review, 52*, 295-304.

Shotel, J., Iano, R., & McGettigan, J. (1972). Teacher attitudes associated with the integration of handicapped children. *Exceptional Children, 38*, 677-683.

Sivage, C. (1980). The organizational context of mainstreaming. *Directive Teacher, 2*, 16-17.

Spodek, B. (1982). What special educators need to know about regular classrooms. *Educational Forum, 44*, 295-307.

Stannard, R. (1976). Mainstreaming: Some basis for caution. In R. O'Donnell & R. Bradford (Eds.), *Mainstreaming: Controversy and consensus* (pp. 57-79). San Rafael, CA: Academic Therapy Publications.

Winzer, M. (1984). Mainstreaming the handicapped child: Attitudes of teachers and nonteachers. *Canadian Journal for Exceptional Children, 1*, 23-26.

Winzer, M. (1987). Mainstreaming exceptional children: Teacher attitudes and the educational climate. *Alberta Journal of Educational Research, 33*, 33-42.

Zey, K. (1981). Before you mainstream. *Special Education in Canada, 55*, 11-12.

The Alberta Journal of Educational Research *Vol. XXXIV, No. 4, December 1988, 403-417*

PERSPECTIVES

Policy Research and Politics

Jane Gaskell
University of British Columbia

This paper explores what it means to do research that is intended to be relevant for policy. It argues that policy research should not be seen as a technical exercise that is politically neutral. The methodology and the conceptual framework the researcher adopts have political implications which must be consciously confronted. Some examples of research are discussed to illustrate the point. The implications for how research might be used in the political process are also discussed.

Policy research, policy studies, policy centers are all becoming more common in faculties of education in Canada and elsewhere. The faculties of education at the University of British Columbia, the University of Calgary, and the University of Western Ontario now offer policy studies in various ways. The journal *Educational Evaluation and Policy Analysis* was started by the American Educational Research Association a decade ago. The *Journal of Educational Policy* began a year ago at the State University of New York at Buffalo, and *Educational Policy* was started a year earlier in Britain.

There are a variety of reasons for the appeal of the policy label. It provides a common focus for educational scholars in different disciplines to come together around interesting educational debates. It offers the possibility of bringing in research money from government agencies that are analyzing and evaluating policy. The new label is especially appealing to foundations departments because it ties the foundations disciplines more closely to the field by stressing their utility for practitioners, thus giving them an importance they may have seemed to lack in the eyes of curriculum specialists and teacher educators.

The appeal of the educational policy rubric also reflects larger social trends. Cutbacks in university funding force departments to justify their existence more actively and encourage the consolidation of departments which then need new, more encompassing names for what they do. The availability of funding for research is coming to depend increasingly on demonstrated links with current policy issues; the Canadian Social Sciences and Humanities Research Council institutes a matching grants policy and designates areas of *strategic* importance. Moreover, educational policy issues are assuming a somewhat higher profile in the political arena. The links between education and economic growth are emphasized, and the

---

Dr. Gaskell is with the Department of Social and Educational Studies and the Centre for Policy Studies in Education.

needs of the *information society* are debated. In the United States, documents such as *A Nation At Risk*, the *Holmes Report*, and the Carnegie Foundation's studies of high schools and colleges have engendered national discussion about education and educational policies. In Canada, the federal government's forum on postsecondary education, the creation of Ontario's skill development ministry, and the B.C. cabinet's willingness to link education and economic development are a few signs that educational policy is beginning to matter to a wider public.

The trend toward policy studies is, of course, not universally popular with educational researchers. Many have no interest in being hustled into policy research. They would rather study issues that arise out of the disciplines or out of their own interests. Some argue that doing policy research is destined to confirm the second-class status of faculties of education by entrenching the distinction between the *applied* research we do and the *fundamental* or *basic* research done elsewhere in the university. Some see policy studies as a fundamentally conservative enterprise where critical scholars sell out in order to get contracts with policy makers. Policy studies can look boring, conservative, and second-rate.

In this climate, it is important to consider more formally some different conceptions of *policy studies* or *policy research* and to encourage a wider discussion of the issues involved. Policy research encompasses both research on how the policy process in education works and research that is designed to have an impact on social, and more particularly educational, policy. The first is an important field of educational research. It is, however, likely to remain a specialized area of study, appealing to those interested in the politics of education and organizational analysis. It is the second kind of policy research—research designed to inform policy—that can potentially encompass most of the research that we do in education, for educational research is linked to the improvement of policy and practice.

What happens if we start to conceive of our research as having implications for educational policy? The nature of the links between the political process and the research process are at the base of the problem. Doing policy research entails doing research that is linked to the political process in some way and that is meant to affect politics. How would this change what we do? Are the changes ones that should be encouraged or discouraged? What are the pitfalls and what are the benefits?

In this article I argue that to do research that is relevant for educational policy does mean shaping one's research in ways that self-consciously take into account the political process and the researcher's role in it. Instead of setting policy research apart from other educational research, this recognition makes explicit what is implicit in all research. There is increasing awareness that research is not a neutral fact-finding expedition that can be independent of political debate. The ivory tower offers no protection from political controversy, nor should it be used by researchers to claim exemption from political responsibility.

Instead of entrenching educational research as second-class research, this recognition brings educational research into the mainstream. Research

in all disciplines is shaped by its political and social context, and this should be recognized explicitly instead of being hidden away behind a veil of scientific objectivity and political neutrality. Research shaped consciously by political agendas is as diverse, as theoretical, and as important for the development of the disciplines as any other research.

Recognition of the implicitly political nature of research also entails rethinking how research might be used in the political arena. Policy research need not be designed to serve only those who hold political office. In a democracy, all sides of an issue deserve a hearing, and research can help many groups articulate and clarify their concerns.

The article begins with a discussion of the traditional view that science and politics should be kept separate from each other. Examples of educational research are used to illustrate the inadequacy of this position. Finally, some lessons are drawn for researchers about more adequate ways to conceptualize the relation between the political process and the research process.

The Separation of Research and Politics

There is a long intellectual tradition that makes a clear distinction between science and politics and argues that intellectuals have a position outside the political arena to pursue, in an *ivory tower* as we have come to call it, truth at all costs. Politics, in this view, interferes with research objectivity and should be avoided. It is this tradition that makes policy research seem suspect and second-rate.

Weber's classic essay on science as a vocation outlines one version of this argument:

> It is said, and I agree, that politics is out of place in the lecture room.... When the docent is concerned with politics, it belongs there least of all. To take a practical political stand is one thing, and to analyze political structures and party positions is another. The true teacher will beware of imposing from the platform any political position upon the student whether it is expressed or suggested. (Gerth & Mills, 1958, p. 146)

Weber was not for a moment against *scientists* being involved in politics as citizens. What he painstakingly argued was that in the process of doing science, political presuppositions must be avoided. Polin gives a more recent version of this argument in Newsome and Buchbinder's (1988) book on university politics:

> Any university worthy of its name has to resist cultural, ideological and political pressures that would bend it to requirements irrelevant to the life of the mind and loyalty to the truth. A host of pressure groups seek to subvert academic ideals and corrupt the university and their methods include propaganda, dogma, indoctrination, and cultural, even physical terrorism. (p. 57)

Politics corrupts the academic's search for truth. It biases conclusions and bends honest inquiry; politics should be kept out of university inquiry and out of research.

The promise of objective knowledge revealed through university research also implies, however, that this research should be well-funded by the politicians. The society needs the knowledge of the academic community.

Politicians should avail themselves of the knowledge of academics in order to solve social problems, for the unbiased research-based knowledge of the university is superior to the political bias and half-truths of the political process. This sums up an argument that has a powerful hold over our conceptions of politics and research.

The tradition of the disinterested researcher who can provide rational solutions to social problems can be traced back to the Enlightenment. Torgeson (1986) describes it as the *first face* of policy analysis. It came to dominate the attitude of policy makers and social scientists in the 1960s and '70s. At a time when the *end of ideology* was heralded for the postindustrial era and more sophisticated and powerful methodologies were developed in the social sciences, the promise of social science knowledge for social engineering was at its height (Callahan & Jennings, 1983).

Its philosophy of science was positivist. Science is contrasted with ideology, and political bias. Science is a matter of facts, not values. Since Kuhn's (1970) discussion of paradigms in science and the barrage of critical work on positivist conceptions of science that followed (Carr & Kemmis, 1986), it has become increasingly hard to sustain this view.

Politically, it encompassed "a dream of putting an end to the strife and confusion of human society in favor of an orderly administration of things based on objective knowledge" (Torgeson, 1986, p. 34). This tradition tends to be profoundly antidemocratic in wanting to substitute the knowledge of experts for the knowledge of ordinary people, even though the researcher can be portrayed as the handmaiden of the politician, the disinterested technical problem solver who is called on by the political process only when answers to technical questions are required.

> Policy research is, as I understand it, research which should be undertaken by qualified researchers in order to produce evidence on issues deemed to be important by policy makers. Such evidence can then be used in the formulation or revision of laws, or decrees, or educational guidelines. There is an implicit belief or hope that decision oriented research can come up with results that can be used for resolving current problems or issues in education. (Postlethwaite, 1986, p. 135)

Policy research in this tradition is characterized by instrumental rationality. Researchers are asked to look at questions like the following: Was Head Start effective? How should we teach reading? Should vocational training take place in the high schools or the colleges? The researcher does a new study, or consolidates evidence from others to provide the best answer available, based on the most systematic inquiry. The researcher is invaluable for policy makers by making the best information available.

The rhetoric and the dream of objective and unbiased research on policy issues persists, partly because of the interests it serves. Policy makers use the notion of objective research as a mantle to cloak and legitimate their decisions. Researchers use their aura as experts to persuade politicians to come to them for advice. The idea of the university researcher as disinterested, politically neutral seeker of the truth has great appeal to the academy as well as to politicians. It places university researchers in the position of being the experts in areas of their research and of being untainted by the

self-interest and bias of those outside the academy. It places them in the best position to know the truth. It offers the politician a place to go for *unbiased* and knowledgeable answers.

Instead of clinging to these inadequate notions which misrepresent what we do to ourselves and to others, we must begin to develop alternative conceptions of how our research might be related to policy. The clear separation between research and politics is impossible to maintain. First, good research is not uninfluenced by political processes. The political location of so-called unbiased researchers has been clearly revealed in any number of political controversies. Developments in the philosophy of science have pointed to the inadequacy of the claim that facts can be separated from theories and values in any research endeavor. A new conception of social science must underpin our efforts to create a new version of how politics and research might come together.

Secondly, the old view of the separation of politics and knowledge is profoundly undemocratic in its implications. Putting such a high value on the knowledge of experts devalues the knowledge of ordinary people and the value of the political process in devising solutions to social problems. A new conception of the uses of research knowledge in the political process must be part of the reconsideration of the relationship between politics and research.

The Impact of Politics on Research

The impact of personal and social positions on educational research is increasingly recognized (Allender, 1986). As Lather (1986b) puts it, "just as there is no neutral education, there is no neutral research.... Research paradigms inherently reflect our beliefs about the world we live in and want to live in" (p. 257).

No research can address the entirety of any issue, much less a controversial and complex one, as any interesting policy issue is. No single viewpoint, however well informed, is adequate to the task of definitively characterizing social reality. Research questions are framed by our interests and experiences and are based in what Lindblom and Cohen (1979) call *ordinary knowledge*, that is, commonsense conceptions of the issue that are abroad in the society and the profession. Researchers look at the world through the spectacles of their conceptual frameworks which are expressed in language. Researchers must simplify and define a frame of reference and a series of questions in order to do their research. They must be clear on the purposes of their research. But in doing what we learn to do as competent researchers, we make choices that have consequences for what we will be able to conclude and how we will be able to participate in the political debate.

It is easy to see this in examples of educational policy research. The Coleman report on equal educational opportunity seemed like a definitive if controversial research study in 1968 (Coleman et al., 1966). But its use of regression analyses to relate student and school characteristics with achievement on standardized tests was based in the theoretical and methodological assumptions of its time. *Effective school* research followed the Coleman report discussion. It pointed out some of the theoretical and policy

pitfalls of using individual and family characteristics instead of school characteristics as independent variables. But it too can be seen as the product of a particular time and place (Cuban, 1984).[1]

Research on the adequacy of private schools provides a good example of the interaction of research and politics. Coleman's recent study comparing private and public schools in the United States has been subject to a great deal of methodological critique that points to the interaction of methodological decisions and political conclusions (Hoffer, Greeley, & Coleman, 1985). Bergen's (1987) article on approaches to research and policy on private schools in Ontario and Alberta shows how government commissioned studies are shaped by their political roots.

Some less well known Canadian research done in British Columbia in the late 1970s provides another interesting example. The research was carried out just after the passage of the controversial Bill 33 which provided funds for independent schools. It was being partly funded by the Ministry of Education and directed by Erickson (1979), an American researcher who was known as an advocate of independent schools. The B.C. Teachers' Federation refused to cooperate. The controversy was reported at length in the newspapers.

Erickson's (1979) research was based on an argument that schools characterized by a sense of community, *Gemeinschaft*, as he called it, were more effective. He had called on a substantial body of literature to argue that in such schools students learned more, students felt better about themselves and their school, and teachers were more motivated and concerned about their students. *Gemeinschaft*, he argued, was necessary for a good learning environment. He went on to design a study that explored different dimensions of *Gemeinschaft* in the schools, and he compared public and private schools using these dimensions. He developed standardized scales designed to measure *Gemeinschaft*, and he found, not surprisingly, that private schools scored higher on the scales he used. He concluded, to oversimplify his argument, that private schools were better schools.

Erickson argued that his research was carefully done and that his involvement in the issue should be seen not as a liability that compromised his judgment, but as an asset that motivated him to do the research in the first place and helped him to clarify the questions. After all, he argued, if we could only do research on issues we did not care about, much of the interest of doing research would disappear. Moreover, scientific canons of inquiry would safeguard the study from bias. Rules about how to use evidence and the fact that the data were freely available for reanalysis ensured his neutral conduct of the research.

There were, of course, questions one could raise about the conduct of the research. Sampling problems, scales that were questionable, and conclusions that went beyond the data meant it was possible to pick methodological holes in the research, as it is possible to pick methodological holes in most research. But the more interesting issue was the way he framed the research in the first place and the kind of analysis that followed from this.

Erickson's questions had to do with creating a sense of community in schools because he valued a sense of community, and he had a hunch there was more of it in private schools than in public ones. He set out to show that what he had observed informally, what led him to prefer private schools, could be demonstrated formally by social science research. And sure enough, he found the differences he was looking for, on the average.

But think for a moment what a study would look like if opponents of private schools asked their questions, if they set out to show the strengths of public schools and the weaknesses of private ones. If they were worried about elitism, they would look at admission standards that discriminate against different kinds of students. If they were concerned about fundamentalist religious schools, they might take a look at the curriculum and see how democratic or close to scientific norms the curriculum in different schools was. If they were committed to the teachers' federation, they might look at the working conditions and salary levels of teachers and argue that the schools with the best conditions for teachers were the best schools. They might look at the accountability of the schools to the general public. They might look at how adequately schools cater to the needs of all students, not simply academic ones, by offering career education programs or programs for students with special needs. Research directed toward assumed weaknesses in the private school would gather different kinds of data, would test different models, would hold up different versions of the public interest, of good schools, of good education.

In short, Erickson poses his question, already thinking from common sense and a look at the literature that it is probably right, and then he engages in a selective search for facts that bear out his argument. This way of going about research is what we must do to develop research hypotheses and look for evidence. But it should be clear what is going on. Inevitably, the views of the researcher do affect the study. To try to do *value free* or *unbiased* research is simply to conceal, or try to conceal, the implicit point of view in the questions that are asked and the search for evidence. Researchers would do better to make clear why the problem is framed as it is, to justify the framework they are using, and to place it among competing concerns with the relevant issue.

When a particular study is located within the political debate, it becomes clearer that there is a range of other research studies that might be done on the same topic reflecting different political agendas. It makes clear that research on any politically controversial topic should be done from a variety of different political positions, not simply on one side of the argument. The limited sources of funding for research often make this difficult. Obviously, some sides of a political argument are going to be better funded than others. The B.C. government and the National Institute of Education provided Erickson with $500,000 to do his study. It is unlikely that the B.C. Teachers' Federation and the B.C. School Trustees would come up with an equal amount of money. However, it is interesting that the Federation moved in the direction of using social science research to make their case when Bill 33 was announced. They commissioned a public opinion poll on selected

questions that they were interested in. They did not ask about *Gemeinschaft*. They did ask about whether schools should be accountable and should follow the core curriculum. They found that the public had attitudes which seemed to be antithetical to the private schools, and they communicated their research to the government in an attempt to influence policy. Here is an example of how research can be used on both sides of an issue as part of a public debate.

This controversy illustrates the political import of research questions and research design. How well the research is carried through is a critical issue and must be subject to public debate and scrutiny. But debate about the adequacy of the data collection must not substitute for debate about the adequacy of the assumptions in which the research is situated. It is these assumptions that must be made visible, debated, and formulated in different ways by different researchers. It is these assumptions that have political implications for what can be discovered and concluded and what can be introduced into the policy debate.

To clarify the different levels at which a piece of research can be judged is critical. Donmeyer (1985) argues that there are three such levels:

> First order mistakes occur when the evidence cannot support propositions that have been framed by using a particular language.... Second order mistakes occur when the language used to frame propositions is not adequate for particular purposes ... third order mistakes relate to the inadequacy of the purposes themselves.

First order mistakes involve good use of evidence, and it is important that seeing the political character of research is not taken to mean shading the evidence to suit one's own presuppositions. The reason one frames the questions in a particular way and carries out the research is to find out as much as possible. Good policy research, like any research, must sift and weigh evidence carefully. As Lather (1986a) has argued, we must develop and refine techniques that ensure the credibility of our data and minimize the distorting effects of personal bias on the logic of evidence. It does not do anyone any good to get information that is not as carefully gathered and weighed as possible (p. 19).

Second order mistakes involve the usefulness of a conceptual framework for the political purposes they are being asked to advance. Here researchers must ask methodological and theoretical questions in light of their political purposes. For example, will interviews or questionnaires answer the questions better, how should questions be asked, should the unit of analysis be individuals or institutions, should students or teachers be questioned? The researcher must clarify how questions of individual agency and structure are being handled and what the methodology is going to illuminate and conceal. These decisions involve knowledge of competing frameworks and conceptual languages and an appreciation of their different implications. The question of what the research is trying to understand—what the political purposes are—should underlie these kinds of decisions.

Third order mistakes involve an analysis of the political purposes themselves. Doing politically located policy research does not mean that there can be no dialogue among different researchers about the adequacy of their

political frameworks and purposes. The relativist view that one person's politics is as good as another's and that differences among research frameworks are simply a matter of personal political preference trivialize the most important questions about our commitments as researchers and educators. To justify one's research choices is to engage in discourse about political questions, about the nature of good education in a good society. This dialogue is a critical one. It brings political discussion to the heart of the academic process.

Political self-consciousness means not just personal confessionals at the beginning of research reports but also a serious conceptual analysis of the frameworks that are being used and an argument for their utility. Politically situating the research in this way can help it have an impact, while at the same time it improves the conceptual logic and academic value of the study. Policy research becomes not a second-class research but research that does what all good research should do, and incidentally what the Social Sciences and Humanities Research Council demands applicants for research grants in all disciplines do, that is, justify their conceptual framework.

The Political Uses of Research

If research in education is to be increasingly directed toward policy controversies, it must become more conscious of its own uses in the political process. The lament of social scientists concerned with current questions of policy has been Why don't the politicians pay more attention to us? (Holdaway, 1986).[2] Why don't they commission more studies when they need answers? Why don't they heed the information that is available? Why do they simply use research when it legitimates decisions already taken?

In the traditional view of politics and research, the fact that policy makers do not use research is difficult to understand. The commonsense knowledge that politicians do bring to their decisions is seen as flawed and often inaccurate. Researchers see themselves as having superior knowledge The scientist who is neutral, is outside politics, has no axe to grind, and can see with the sophisticated lenses of social science can clearly offer guidance—but his or her expertise is ignored.

Within the *first face* of policy research, the problem of the nonutilization of research findings can be treated as a technical one of bringing together researchers and policy makers. Some look at the lack of avenues of communication between the two communities. Policy makers do not read academic journals, and academics rarely use other media for communicating their ideas. The differences in time frame between academic researchers and policy makers is also recognized as a problem. If a researcher takes two years to do the research, and the policy decision must be taken in two weeks, no wonder the academic research is not used.

The problem can also be traced to the motivations of policy makers. The costs of the social scientist's good recommendations may be too high in terms of dollars or political support. The concerns of policy makers may be not the public welfare, but their own political longevity. "The policy-making value of any study depends on whether its recommendations can be imple-

mented while at the same time enhancing the political strength and social status of the government" (Bergen, 1987, p. 304).

But once the political assumptions built into the research are recognized for what they are, other explanations for why research is not used by policy makers become apparent. The problem is not simply a technical one of making the research known and available to those who should use it. The problem is the congruence between the political frameworks of the researcher and the political actor. If the political framework of the researcher does not fit with that of the politician, it is unlikely that the research will be perceived as useful. If the research does not address the concerns of someone who is involved in the determination of policy, it will not be relevant to the debate around the policy. To be useful, policy researchers must not only abandon their quest for political neutrality but must also take into account explicitly which audiences might use research in a political arena. To have their research used, researchers need to recognize that policy making is properly a political process and to locate themselves within this political discourse rather than outside it.

This argument has tended to produce a rather conservative face for policy research. Policy research, it is argued, must take into account the way the government sees things. Policy researchers must learn to frame their research from the point of view of those who will make the decisions. May (1986) urges more attention to the political feasibility of policy options. Bergen (1987) writes that "The policy making usefulness of a study is contingent on the degree to which its recommendations reflect the reality of economic, social and political forces which sway the decisions of governments" (p. 304).

Coleman and Laroque (1983), recognizing the necessity for taking into account the actual political process in the design of policy research, argue that researchers need to pay more attention to what politicians want from researchers when they design their research. They suggest policy research should

> investigate problems which are of concern to them, and should be designed to provide data and interpretations which have predictive validity concerning the probable consequences of policy options.... [It] should concern itself with those policy options which are important to policy makers [by] providing relevant and reliable predictive generalizations. (p. 247)

In this mode, the researcher remains a technical expert in the service of the administration by adopting the current discourse and a scientific mantle and by finding technical answers. Administrators are assumed to be benevolent, interested in the welfare of all, and open to advice from researchers. The role of the researcher is restricted to providing answers to administrator's questions. To tailor research to their needs is to serve the collectivity.

Self-consciously building the assumptions of the administrator into the formulation of the research problem recognizes that research must have an audience and recommends adopting a framework that takes politics into account in order to reach the intended audience. But the particular political

assumptions that are suggested are just what scare so many away from a policy framework and result in the accusation that policy research sells out to the highest bidder. If policy research involves serving those with power and identifying politicians and administrators as benevolent and powerful, critical research is imperiled. The options considered by administrators will fall within a narrow range, and too many limitations will be placed on the discussion of issues. If policy research means limiting oneself to considering viable options for those with power, what happens to challenges to the existing political order, to the lively debate that has characterized conceptually innovative research? Policy research that accepts these limitations reinforces the structure and ideology of the existing world.

It is important to broaden our ideas of the audience for policy research beyond administrators in schools and politicians who want to retain office. We need a new version of the political process to work with, one that encourages democratic decision making instead of administrative fiat, one that understands the complexity of arriving at policy outcomes instead of positing a rationalist process.

If policy research is conceived of as addressing a democratic polity with conflicting interest groups and shifting agendas instead of addressing the benevolent administrator with a defined problem, much broader uses for research in the political process open up. Researchers can explore issues from many potentially conflicting points of view including that of the student, the parent, the teachers' federation, or the government agency, to mention only a few. It can clarify, legitimate, and expand the political agendas these groups start with instead of simply solving problems for them. The researcher can work in dialogue with the groups whose concerns he or she is addressing. What matters is that the research is located, contextualized, and addressed to a specific audience.

Research can serve to highlight new questions and problems, to help frame the political discussion, as well as to solve problems. As Worth (1978) puts it,

> We must start with the recognition that the problem-solving image of policy formation is probably a myth.... The functions of policy research ought to include the identification of problems as well as the analysis of possible solutions and the evaluation of outcomes. (p. 5)

Weiss (1977, 1983) has stressed what she calls the *enlightenment* functions of research:

> Not single findings, one by one, but ideas from social science research appear to affect the development of the policy agenda. Social science research is one of the sources from which participants in the policy process derive their sense of how the world works. (1983, p. 219)

Research in education can articulate educational problems in ways that are official and public. Research on drop-out rates, literacy levels among Native people, or the effects of streaming on high school graduation can be a political statement on behalf of groups who have been excluded. Torgeson (1986) uses the Berger Inquiry as an example of policy research that manages to give a voice to those who would not normally be heard. The

government funded a wide variety of groups to do research and put together briefs that could then be considered in the inquiry about the Mackenzie Valley Pipeline.

> The Inquiry solicited and received scholarly reports on various dimensions of the pipeline issue. The research, however, was not simply accepted at face value. Impact statements countered one another; study challenged study. With replies, rebuttals, cross examinations, debate moved at a high level of sophistication. (Torgeson, 1986, p. 47)

Educational policy makers and educational researchers can learn from this example. Instead of calling for definitive studies of educational problems, we need to encourage diverse constituencies to take the research enterprise seriously. In the process, the public can come to see the value of the research enterprise, and our research can be informed by their understandings.

If an issue is important enough to research, it is important enough to introduce into the political process. Researchers need to see themselves as advocates of their research and their findings. They can be advocates for children, for equity, or for teachers if their research is framed in ways that allow it. Researchers need to follow through on the political commitments they make in their work. The simple separation of politics and research is no longer an adequate conception of how we work, especially in professional faculties committed to the improvement of practice.

The debates around Erickson's research on private schools, which I discussed earlier in this paper, provide a concrete example of the mismatch between research and political agendas. Erickson's concerns for the effectiveness and sense of community in private schools were far from the concerns of policy makers at the time (Downey, 1979). Some of the issues that did matter were issues of principle—*the right to choose*; some were issues of self interest—the power of organized religion. The question of where votes would be won or lost mattered a great deal to the government. While Erikson's research dealt with an important educational issue, with the quality of schooling, with, at least by inference, effects on children, the policy debate did not appear to be much influenced by these issues. This is too bad. One role of a controversial study is to highlight issues of concern to children and introduce them into the debate.

But introducing a new issue is difficult unless there is a political constituency that will take it up. If the assumptions of the researcher are shared by some political actors, the research is more likely to be taken up and used. In most policy debates, perceptions and vested interests are strongly held. Locating one's work within the existing debate, even if it ultimately goes beyond this debate, is often a good idea.

Erikson's (1979) work did not change many minds about whether private schools are good schools, bad schools, or, as is more likely, a mixed bag. People's commonsense understanding of the world is powerful, and they are reluctant to give it up in the face of *scientific* evidence often gained from a questionnaire on an odd sample of schools. There are many examples of this phenomenon in education. Perhaps the best known is the

issue of class size. Research shows that students do not learn any more in small classes than they do in larger classes, other things being equal. However, teachers are convinced that students do learn more in smaller classes, and research evidence will not convince one side or the other, because each operates with different ideas of what counts as good schooling.

Consideration of the existing debates is likely to inform the political decisions I have argued every researcher must make. My own research has made this clear for me. In examining questions of equality for women, the positions of liberal feminists, socialist feminists, and radical feminists all need to be considered in deciding how to frame questions of equity (Gaskell & McLaren, 1987). If equality of access to existing programs is the issue, one looks at barriers to entry for women in nontraditional programs. If equal pay for work of equal value is the issue, one looks at how the value of women's traditional work is defined. How to frame important research questions grows out of a consideration of important political issues.

Where does this leave the researcher in relation to the political process? Not in the position of philosopher queen or king, sitting in the best position to decide on policy outcomes but rather in the position of a well-informed political actor engaging in a democratic political process where argument, evidence, and political power matter. What research can do is test out particular views, modify and adjust them a bit in light of the real world, and articulate them more clearly, more forcefully, more consistently. It can also, in the long run I think, improve the nature of the debate by adding information and clarity to the issue.

Conclusion

I have argued that research is an inherently political process reflecting the values, the concerns, the commonsense—socially and historically located—assumptions of the researcher. To recognize that research is involved in political choices is not to argue for any particular theoretical or methodological position. It is certainly not to disrespect the particular expertise that is necessary for good research. It is, however, to suggest that policy researchers should see themselves as political actors and work to have their research reflected in political debates. It is not to argue that research should serve those in powerful positions, but it does mean that research should seek out an audience and be located in terms of the concerns of this audience.

This involves both humility and daring on the part of the researcher—humility in recognizing that the way we frame problems may not be the way others, including policy makers, do. We have no exclusive claim on right answers because we have done research. Others have every right to frame questions their way, and we can often learn if we listen to them. As researchers, we do not have the institutional responsibility or the knowledge necessary for making policy decisions. But recognizing the political nature of research also involves daring. Researchers cannot hide behind their technical role, arguing that all the political choices are made elsewhere. In recognizing the way carrying out research involves political decisions, re-

searchers must be prepared to defend their political positions and to take this defense forward into the political arena where decisions are made.

Politicization, then, is not a problem. Good policy research should be firmly situated in the literature, in policy discussion, in relation to the researcher's purposes and audience. Rather than claiming neutrality, every project should expose the commonsense underpinnings that situate it in a political discourse. Each part of that discourse has a place for good research.

Acknowledgments

This paper was supported by funds from SSHRC grant No. 410-85-0356.
I would also like to thank Joel Weiss and the AJER reviewers for reading and commenting on the manuscript.

Notes

1. See Wexler's work (1976, 1987) on the sociology of education for a detailed analysis of the social and political conditions that affected its direction and content.
2. Worth (1986) sounds a more skeptical note when he writes that educational research is such a poor guide to action in the classroom or the boardroom that we ought to feel relief, rather than regret, that it is so seldom used. Otherwise, researchers might be the scapegoats of our educational system and educational research might enjoy even less support than it does now. (p. 175)
   See Chester Finn (1988) on the same point.

References

Allender, J.S. (1986). Educational research: A personal and social process. *Review of Educational Research, 56*(2), 173-193.

Bergen, J.J. (1987). Government sponsored private school studies in Alberta and Ontario. *Alberta Journal of Educational Research, 33*(4), 292-306.

Callahan, D., & Jennings, B. (1983). *Ethics, the social sciences, and policy analysis.* New York: Plenum Press.

Carr, W., & Kemmis, S. (1986). *Becoming critical: Education, knowledge and action research.* London: Falmer Press.

Coleman, J., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., & York, R.L. (1966). *Equality of educational opportunity.* Washington, DC: US Government Printing Office.

Coleman, P., & Laroque, L. (1983). Linking educational research and educational policy via policy-relevant research. *Alberta Journal of Educational Research, 29*(3), 242-255.

Cuban, L. (1984). Transforming the frog into a prince: Effective schools research, policy and practice at the school level. *Harvard Educational Review, 54*(2), 129-151.

Donmeyer, R. (1985). The rescue from relativism: Two failed attempts and an alternative strategy. *Educational Researcher, 14*(10), 13-20.

Downey, L. (1979). *The anatomy of a policy decision.* Vancouver: University of British Columbia, Faculty of Education.

Erickson, D. (1979). *The funding of independent schools.* Research report submitted to the Ministry of Education, Victoria, BC.

Finn, C.E. (1988). What ails education research. *Educational Researcher, 17*(1), 5-8.

Gaskell, J., & McLaren, A. (1987). *Women and education: Canadian perspectives.* Calgary: Detselig.

Gerth, H.H., & Mills, C.W. (1958). *From Max Weber: Essays on sociology.* New York: Oxford University Press.

Hoffer, T., Greeley, A., & Coleman, J.S. (1985). Achievement growth in public and Catholic schools. *Sociology of Education, 58*(2), 74-97.

Holdaway, E.A. (1986). Making research matter. *Alberta Journal of Educational Research, 32*(3), 249-264.

Kuhn, T. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago.

Lather, P. (1986a). Issues of validity in openly ideological research: Between a rock and a soft place. *Interchange, 17*(4), 63-84.

Lather, P. (1986b). Research as praxis. *Harvard Educational Review, 56*(3), 257-277.

Lindblom, C., & Cohen, D. (1979). *Usable knowledge: Social science and social problem solving*. New Haven: Yale University Press.

May, P.J. (1986). Politics and policy analysis. *Political Science Quarterly, 101*(1), 109-125.

Newson, J., & Buchbinder, H. (1988). *The university means business*. Toronto: Garamond Press.

Postlethwaite, N. (1986). Policy-oriented research in education. *Oxford Review of Education, 12*(2), 135-151.

Torgeson, D. (1986). Between knowledge and politics: Three faces of policy analysis. *Policy Sciences, 19,* 33-59.

Weiss, C. (1977). *Using social research in public policy making*. Lexington, MA: Heath.

Weiss, C. (1983). Ideology, interests and information: The basis of policy positions. In D. Callahan & B. Jennings (Eds.), *Ethics, the social sciences and policy analysis* (pp. 213-248). New York: Plenum.

Wexler, P. (1976). *The sociology of education: Beyond equality*. Indianapolis: Bobbs Merrill.

Wexler, P. (1987). *Social analysis of education: After the new sociology*. London: Routledge & Kegan Paul.

Worth, W.H. (1978). Perspectives on policy formation. *Canadian Journal of Higher Education, 8*(2), 1-7.

Worth, W.H. (1986). Research impotence: Blessing or blight? *Alberta Journal of Educational Research, 32,* 175.

The Alberta Journal of Educational Research *Vol. XXXIV, No. 4, December 1988, 418-420*

BOOK REVIEW

INTRODUCTION TO THE ECONOMICS OF CANADIAN EDUCATION.

By Donald M. Richards and Eugene W. Ratsoy. Calgary, AB: Detselig Enterprises, 1987, 127pp.

This volume is a welcome addition to the writings on educational economics and finance in Canada. It is designed to provide a basic perspective on economics to university students and policy makers. I was pleased to receive a draft copy and offer a suggestion several months ago. Consequently, this review provides an opportunity to see if the authors acted on my advice. Let me take a walk through the book and then present some general reactions.

Chapters 1 and 2 cover basic concepts from economics which the authors believe useful for those whose background in the subject is scant. I would concur that this is a good idea because most prospective readers are unlikely to be acquainted with the nomenclature. Chapter 1 is presented in the raw, however. While clearly laid out in simply-stated language, it could have been made less dry with the use of examples from education. References to commodities are fine, but educators are in the service business. The next chapter, a well-stated account of wealth and economic activity, is accurate to the modest knowledge of this reviewer. There is also more information on Canada here, although more stress on business cycle indicators such as the consumer price indexes and interest rates might have been helpful. Chapter 3 extends the discussion to national economic policy and is mostly devoted to the Economic Council of Canada's annual reviews from 1964 to 1971. A useful summary section provides five main goals for the Canadian economy. One wonders why the descriptions are restricted to those eight reports, particularly as the emphasis on government intervention has been reduced so much since that era.

The next three chapters focus on investment in education. Chapter 4 presents the main concepts such as consumption, spillovers, and public versus private goods. However, no sources for the ideas are given so that a reader is given no leads to pursue other references. A section on techniques in working with compound interest lays out the formulas, but except for one example not applied to education, it leaves the reader to generalize on his or her own. Chapter 5 brings a number of welcome Canadian studies showing returns on investment in education, all but two being at the postsecondary level. They could serve as fine building blocks for future research. Just why the selected studies were all completed during 1965-70 (except for one in 1980) is not clear. It might have been useful to know about educational returns on investment during the last two decades, particularly because wage rates for unionized labor may have boosted lifetime earnings for some with fewer years of education. Chapter 6 grapples with some of the different approaches used with human capital studies. However, the writing is strict-

ly authoritative, except for references to Dennison. No other connection is made to the literature.

The next three chapters address costs, revenues, and governance. Chapter 7 illustrates the dramatic increase in the costs of education well. It presents approaches to the studies on costs and connects them with Canadian sources. The thorny problems of revenue generation are tackled in Chapter 8. Again, one wonders where certain principles such as those of taxation (p. 92) come from. However, the main ideas are clearly stated with balanced coverage. Chapter 9 hits a number of economic policy issues facing educators. These are so important that I would have preferred more stress on them.

The last quarter of the book is devoted to the smaller-scale issues surrounding the production of educational services. Chapter 10 describes the various factors, processes, and outputs which characterize the general production model. The next chapter then looks at different ways in which the model may be conceived and applied —perhaps the best chapter in the book. It includes work on effective schools which demonstrates the willingness of the authors to conceive of the economics of education rather broadly. The final chapter takes a look at production function studies (broadly defined) and offers some of the standard criticisms of them.

Here are three general reactions. One is the need for more clearly documented sources, particularly for students. While policy makers may not need to know where ideas come from, a book used as a text for courses in educational economics and finance requires that ideas be connected to prior works. (For those interested, this was the recommendation transmitted earlier.) Another problem is the generally descriptive nature of the volume. While some descriptive material is necessary, I think it would have been useful for the authors to have provided many more comparisons and contrasts of the ideas as they were presented. While the onus is always on the reader to make the connections and add the criticisms, this reviewer believes that for both students and policy makers, the volume would have been strengthened with a greater amount of analysis than it now contains. A third general reaction is that I would like to have seen a thirteenth chapter which spelled out some of the exciting policy choices derived from the earlier chapters. Such a chapter would have demonstrated the power of the ideas more completely. While many of these issues are addressed in passing, a book like this could have provided a sense of potential Canadian policy directions for the next decade.

It would be easy to dismiss the economics of education. The concepts are so compelling, but the data are so damnable. Yet the ideas and facts brought forward in this book are valuable ones. Richards and Ratsoy have given us a highly readable handbook of interest to those who have a concern for educational policy, particularly at the provincial and local levels. What is its use as a text? It should serve well as a supplementary volume, though instructors will probably prefer to remain with old standbys such as Benson or Cohn for their greater breadth and depth.

What is the bottom line? Richards and Ratsoy are to be commended for their willingness to produce this volume which integrates the *econ. of ed.* into the Canadian context. This work has obvious utility for students and policy makers because it makes the right assumptions about the level of knowledge and issues of key interest.

Daniel J. Brown
Faculty of Education
University of British Columbia

THE UNIVERSITY OF CALGARY

THE JOURNAL OF EDUCATIONAL THOUGHT

Published three times yearly by the Faculty of Education
The University of Calgary

Telephone (403) 220-5629

The Journal of Educational Thought promotes speculative, critical and historical research concerning the theory and practice of education in a variety of areas: administration, comparative education, curriculum, educational communications, evaluation, instructional methodology, intercultural education, philosophy, psychology, and sociology. Critical and controversial views on the field of education or on previously published articles in The Journal of Educational Thought appear in a section entitled The Forum, which also contains papers of a less formal nature. The book section includes short book reviews as well as extended book review essays. The Journal is international in scope, qualitative in nature, and serves a broad readership; specialists in the areas mentioned, scholars, and the public in general.

The Journal is indexed in *Canadian Index* and by the *Educational Resources Information Centre (ERIC)*. Abstracts of appropriate articles appear in *Educational Administration Abstracts, Language and Language Behaviour Abstracts*, and *Sociology of Education Abstracts*. Available on microfilm from University Microfilm, Inc., Ann Arbor, Michigan.

Subscription Rates:

Effective January 1, 1986

$22.00 - 1 year (individual)
$27.00 - 1 year (institutional)
$ 9.00 - single issue
$ 4.00 - back issue
$ 7.00 - special issue

The Journal of Educational Thought
Faculty of Education
The University of Calgary
2500 University Drive N.W.
Calgary, Alberta, Canada
T2N 1N4

Instructions for the preparation of manuscripts appear on the last page of each issue.

Index, Volume XXXIV, 1988

Anderson, R.N., Greene, M.L., & Loewen, P.S. *Relationships among teachers' and students' thinking skills, sense of efficacy, and student achievement.* No. 2, June, 148-165.

Arnove, R.F., & Graff, H.J. *What have national literacy campaigns taught us?* No. 3, September, 215-223.

Bacchus, M.K., & Torres, C.A. *Framework for a comparative study of adult education policy implementation in Mexico, Tanzania, and Alberta.* No. 3, September, 320-329.

Carney, R.J. *The evolution of Indian adult education in Alberta.* No. 3, September, 294-309.

Clifton, R.A., & Roberts, L.W. *Social psychological dispositions and academic achievement of Inuit and non-Inuit students.* No. 4, December, 332-343.

Das, J.P. *Intelligence: A view from neuropsychology.* No. 1, March, 76-82.

Duquette, C., & O'Reilly, R.R. *Perceived attributes of mainstreaming, principal change strategy, and teacher attitudes toward mainstreaming.* No. 4, December, 390-402.

Eckler, J., & Weininger, O. *Play and cognitive development in preschoolers: A critical review.* No. 2, June, 179-193.

Ewanyshyn, E.L., & Konrad, A.G. *Perceptions of control over educational decisions in Alberta.* No. 2, June, 134-147.

Fagan, W.T. *Literacy in Canada: A critique of the Southam Report.* No. 3, September, 224-231.

Fris, J., & Balderson, J.H. *Leaders' priorities and the congruity imperative.* No. 4, December, 375-389.

Gaskell, J. *Policy research and politics.* No. 4, December, 403-417.

Kobasigawa, A., Chouinard, M., & Dufresne, A. *Characteristics of study notes prepared by elementary school children: Relevancy and efficiency.* No. 1, March, 18-29.

La Belle, T.J. *Nonformal education for industrialization in Latin America: The human capital approach.* No. 3, September, 203-214.

Latapí, P. *Participatory research: A new research paradigm?* No. 3, September, 310-319.

Malicky, G.V., & Norman, C.A. *Reading processes subgroups in a clinical population.* No. 4, December, 344-354.

Marjoribanks, K. *Cognitive and environmental correlates of adolescents' achievement ambitions: Family-group differences.* No. 2, June, 166-178.

Montgomerie, T.C., McIntosh, R.G., & Mattson, N.A. *Old shadows, new images: A revisionist view of the principal's role.* No. 2, June, 109-133.

Nagy, P., Evans, P.J.A., & Robinson, F.G. *Exploratory analysis of disagreement among holistic essay scores.* No. 4, December, 355-374.

Pannu, R.S. *Adult education, economy, and state in Canada.* No. 3, September, 232-245.

Payne, C.H. *Parents and control of elementary school principal work behavior.* No. 1, March, 2-17.

Robinson, N., & Wood, M. *Why school board members choose to seek or not seek reelection: A test of political efficacy and trust theory.* No. 1, March, 30-41.

Schmelkes, S., & Narro, L. *Adult education and training in Mexico.* No. 3, September, 246-258.

Small, J.M. *Preparation of adult educators in Alberta.* No. 3, September, 287-293.

Small, J.M., & Young, J.H. *Teacher's motivations for participating in curriculum development committees.* No. 1, March, 42-56.

Sumra, S., & Bwatwa, Y. *Adult education, literacy training, and skill upgrading in Tanzania.* No. 3, September, 259-268.

Torres, C.A. *An analytical framework for adult education in Alberta.* No. 3, September, 269-286.

Watson, K. *The crisis of confidence in British higher education.* No. 1, March, 57-75.

Werner, W. *Program implementation and experienced time.* No. 2, June, 90-108.

For Contributors

Please prepare manuscripts in accordance with these guidelines.

Format

- Manuscript in triplicate.
- Typewritten, double-spaced throughout.
- Abstract of approximately 100 words, on a separate sheet.
- Manuscript not to exceed 6,500 words excluding graphics.
- Author's name and affiliation on a separate title page; title only on first page of manuscript to ensure anonymity in the review process.

Referencing

- Sources in parentheses after each reference (direct or otherwise), giving author's name (unless mentioned in text), year of publication, and page number(s).
- Direct quotes of 40 words or less in double quotation marks incorporated into text; those longer than 40 words in indented block format.
- All sources listed alphabetically at the end of the manuscript under the heading *References* using the APA style.
- Explanatory notes, numbered consecutively and marked in the text with superscript numerals, may appear before the References under the heading *Notes*; any citations in notes follow the same format as other references.

Graphics

- Each table and figure on a separate page at the end of the manuscript.
- Figures, charts, and diagrams in camera-ready form.
- Number tables in Arabic numerals and center *Table 1* over each heading.
- Number figures in Arabic numerals before the caption at foot of each graph, e.g., *Figure 1*: ...
- Indicate placement of figures and tables in text, i.e., *Insert Table 3 here.*

Style

- *AJER*'s editorial style conforms closely to the *Publication Manual of the American Psychological Association.*
- Editorial changes may be made to manuscripts.
- For spelling, consult *Webster's New Collegiate Dictionary*; spelling in quoted material remains as in the original.

Diskettes

- If manuscript is produced on an IBM, IBM compatible, or Macintosh computer, please also send diskette. Microsoft Word, MultiMate, WordPerfect, WordStar, Xerox Writer, XyWrite, and DisplayWrite 3 are acceptable.

Submission to *AJER* implies the manuscript is not being considered for publication elsewhere.

AJER is indexed in the *Alberta Education Index, Canadian Education Index, Current Contents/Social and Behavioral Sciences, ERIC/Current Index to Journals in Education,* and *Social Science Citation Index;* appropriate articles are abstracted in *Cognitive Development Abstracts, Educational Administration Abstracts, Educational Technology Abstracts, Family Abstracts, Language Behavior Abstracts, Multicultural Education Abstracts, Psychological Abstracts, Research into Higher Education Abstracts, School Organization and Management Abstracts,* and *Sociology of Education Abstracts,* and reviewed by the *Journal for Research in Mathematics Education.*